5-5-2015

The Impact of Rater Personality and Purpose of Appraisal on Performance Ratings

Matthew D. Marmet
matthew.marmet@gmail.com

The Impact of Rater Personality and Purpose of Appraisal on Performance Ratings

Matthew David Marmet, PhD

University of Connecticut, 2015

Murphy and Cleveland's (1995) four-component model of performance appraisal challenges views taken in past literature on the performance appraisal process by portraying it as goal-directed behavior on the part of the rater. This goal-directed behavior is influenced by several factors including rater characteristics and rating context. To my knowledge, no study has assessed the combined impact of rater personality and purpose of appraisal on performance ratings. By utilizing both Murphy and Cleveland's model and the Socioanalytic Theory of personality (Hogan and Shelton, 1998), the main goal of this study is to uncover a deeper connection between these variables. To assess these relationships, actual developmental and administrative ratings were obtained from 138 managers working in a global consumer products company. Anticipated personality trends in rating behavior are presented based on the Socioanalytic Theory, and it is hypothesized that rater personality will interact with rating context to influence the ratings managers provide to their subordinates. A one-with-many dyadic analysis was conducted to assess these relationships, and partial support was garnered for the hypotheses presented. Although there were no significant main effects of rater personality found, rater Adjustment, Ambition, Sociability, and Learning Approach all significantly interacted with purpose of appraisal to impact ratings in the two settings.

The Impact of Rater Personality and Purpose of Appraisal on Performance Ratings

Matthew David Marmet

B.S., Union College, **2006**

M.A., University of Connecticut, **2008**

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

The Impact of Rater Personality and Purpose of Appraisal on Performance Ratings

Presented by

Matthew David Marmet, B.S., M.A.

Major Advisor_____
R. James Holzworth

Associate Advisor_____
Janet L. Barnes-Farrell

Associate Advisor_____
Allan Church

University of Connecticut
2015

Contents

Introduction 1
Rater Characteristics 4
Rater Demographics 4
Other Rater Characteristics 7
A Four-Component Model of Performance Appraisal 13
Rater Personality 18
The Socioanalytic Theory of Personality 21
The Hogan Personality Inventory 24
The HPI Scales 25
Rating Context 28
Method 41
Participants 41
Procedure 41
Measures 42
Analysis and Results 43
Descriptive Statistics 44
Factor Analysis 44
One-With-Many Design 46
Intraclass Correlation (ICC(1)) Analysis 47
Multilevel Analysis 48
Results of Hypothesis Testing 50
Discussion 57
Interpretation of Results – Main Effects 59
Interpretation of Results – Interactions 59
Limitations 63
Implications 65
Directions for Future Research 66
References 71
Tables and Figures 83
Appendices 105

The Impact of Rater Personality and Purpose of Appraisal on Performance Ratings

Introduction

Measurement of performance in industrial settings has occupied the attention of psychologists for decades because performance prediction, as well as description (rating), plays an important role in all personnel decisions. For instance, criteria are necessary for validation studies and training evaluation, indices of effectiveness and relative worth are necessary for administrative decision making with respect to current employees, and performance-related information is necessary for feedback and employee counseling (Jefferson, 2010; Landy, Barnes, & Murphy 1978).

"Performance appraisal was a term once associated with a rather basic process involving a line manager completing an annual report on a subordinate's performance and (usually, but not always) discussing it with him or her in an appraisal interview" (Fletcher, 2001, p. 473). Unlike this description provided by Fletcher (2001), painting the picture of performance appraisal today is not such an easy task. Although the importance of the process may be common knowledge to most academicians and practitioners in the field, realizing its importance and actually measuring performance accurately are two very separate matters (Landy & Farr, 1980).

Traditional research on performance appraisal asserts that it is a process plagued with errors such as leniency and halo. Barrett (1966) considered leniency (or rating elevation) to be endemic to nearly every rating program. Also, in conducting three separate studies in actual performance appraisal settings, Kane, Bernardin, Villanova, & Peyrefitte (1995) obtained such similar results that they felt justified in classifying leniency as a stable rating characteristic. So as not to lose the significance of halo in this discussion, Cooper (1981) suggests that halo is

likely to be present in every type of rating instrument and Borman (1975) asserts that halo is perhaps the most pervasive type of rater error.

Consequences of these errors have been shown to be severe, and include, but are not limited to, reducing the size and availability of funds to recognize and reward performance, making it difficult to substantiate termination decisions (Bernardin, Cooke, & Villanova, 2000), and making it impossible to portray an individual's performance on different dimensions of behavior (Borman, 1975). With these known negative consequences, the fundamental question that needs asking is "What are the factors that have the potential to influence performance ratings?," as one cannot hope to fix a problem without first understanding it.

The performance appraisal literature has shed light on many factors thought to lead to these types of errors. Included in these are the instrumentation involved in making the ratings (Bendig, 1955; Bernardin, Alvares, & Cranny, 1976; Sharon & Bartlett, 1969), characteristics of the ratee (Butler & Skattebo, 2004; Cleveland & Landy, 1981; Ellis, Ilgen, & Hollenbeck, 2006; Kraiger & Ford, 1985), aids to the rater such as training (Crow, 1957; Gorman & Rentsch, 2009; Woehr, 1992; 1994) and diary keeping (Bernardin & Walter, 1977; Denisi, Robbins, & Cafferty, 1989; Murphy & Balzer, 1986; Raymark, Skowronski, Bevard, and Hamann, 2001), the rating context (Fuqua, Johnson, Newman, Anderson, & Gade, 1984; Lievens, Conway, & De Corte, 2008; Varma, Pichler, & Srinivas, 2005), and characteristics of the raters themselves (Griffeth & Bedeian, 1989; Sackett & DuBois, 1991; Sinclair, 1988).

The main goal of the current research project is to provide an in depth view of the role of a specific rater characteristic, namely personality, in the rating of performance within two distinct organizational contexts: developmental and administrative rating settings. To my

knowledge, there is no study to date that assesses how the interaction of rater personality and purpose of appraisal influences ratings. Do personality effects on rating behavior remain consistent when the ratings are being made for different reasons? In doing so, this project attempts to pursue a view of performance appraisal laid out by Murphy and Cleveland (1995), which portrays the process as one characterized by goal-directed behavior on the part of the rater. This goal-directed view of appraisal ties in nicely with the Socioanalytic Theory of personality (Hogan & Shelton, 1998), in which a person's identity, or the person they think they are and that they hope others will perceive them as, is said to dictate their outward behavior (this will be explained in greater detail below). By partnering Murphy and Cleveland's model with this theory of personality, it is hoped that a better understanding of the influences of rater personality and rating context will emerge, offering a more succinct way for both academicians and practitioners to look at these factors of performance appraisal.

In order to accomplish this goal, existing research in the field of performance appraisal relating to rater characteristics will be discussed. Attention will first be given to simple rater demographics such as race, age, education and sex. After this, a discussion of more complex rater characteristics such as mood, attitude, cognitive processing, and accountability/anonymity will be presented. In doing so, it is hoped the reader will gain an appreciation for the importance given to the rater within the overall process. Following this, Murphy and Cleveland's (1995) model of performance appraisal within organizations will be laid out, providing the spark for an alternate view of the raters' role and the "errors" they make when evaluating performance. Rater personality is shown to be related to this model through the Socioanalytic Theory. The focus will then move to past research on contextual factors such as the politics within an organization, the organizational climate, environments consisting of a multi-source rating context, and the

purpose for which the appraisals are being made. Similar to the sections on rater characteristics, the goal here is to relay the importance of contextual factors in general, and the importance of purpose of appraisal especially. As was briefly mentioned, there are two distinct purposes (or rating contexts) that will be addressed. The first is "developmental" in nature, whereas the second is considered "administrative" (both of which will be discussed in further detail later). The final sections of this particular study will include descriptions of the data collection process, analyses, results, implications for practice, and suggestions for future research.

Rater Characteristics

As was mentioned before, the performance appraisal process is a complicated one that is influenced by numerous factors, one of which is the rater, who is responsible for actually evaluating performance. Put simply, the rater can be considered the user of the instrument in a performance appraisal system (Feldman, 1981). Considerable amounts of past research have investigated the individual differences that make these users unique, and the influence these differences can have on the ratings they assign. Rater demographics, along with a host of other rater-related variables are addressed in the following sections.

Rater Demographics

Past research has shown that a host of rater demographics have the potential to influence the performance ratings they provide to others. Rater race, age, education, and sex have seen attention from researchers, and the results of their work will be discussed below.

Rater Race. Studies involving rater race have shown mixed results in the past. In an investigation conducted in a military setting, Pulakos, White, Oppler, and Borman (1989) assessed the impact of several variables on performance ratings, one of which was rater race.

Although they did find a significant main effect for race in this instance, it was probably due to their large sample size. A population of 8,642 enlisted personnel partook in the study, but the amount of variance in ratings accounted for by rater race was extremely small (less than 1%; Pulakos et al., 1989). Similar negligible findings appeared in earlier research presented by Schmidt and Johnson (1973). The authors found no significant differences among black and white raters, regardless of the race of the ratee.

Sackett and DuBois (1991) compared data from three different sources in an attempt to shed light on the relationship between rater race and performance evaluations. Analysis of a military sample, a civilian sample, and data obtained from a meta-analysis revealed that both black and white raters assigned higher ratings to white ratees. The magnitude of this difference though was much greater for white raters.

In another study conducted by Mount, Sytsma, Hazucha, and Holt (1997), the authors investigated ratings from peers, subordinates, and bosses in a developmental feedback program, and were interested in the impact of race on these ratings. Results indicated that for all perspectives mentioned above, black raters assigned higher ratings to black ratees. Ratings from white raters, on the other hand, differed by perspective. White bosses rated white ratees higher than black ratees, but white subordinates did not. Overall, both white and black managers received higher ratings from black raters (Mount et al., 1997).

Rater Age. Rater age has also been shown to impact performance ratings. A study conducted in a field setting, in which actual performance was observed and assessed, showed that younger raters gave significantly lower evaluations than their older counterparts (Griffeth & Bedeian, 1989). These results however did not match up with those from an earlier study by

Cleveland and Landy (1981). In an attempt to remedy some methodological issues from research they reviewed (i.e., paper people, small sample sizes), these authors sampled 513 managers from a manufacturing organization, and found that younger raters actually gave significantly higher ratings than the older raters.

Rater Education. In terms of yet another demographic variable, rater education level, little research has been done regarding its influence on performance ratings. However, Cascio and Valenzi (1977) did assess this relationship and found that raters with lower levels of education provided higher mean ratings. The authors make a note to mention that although the differences found were statistically significant, their weakness prevents one from making any firm conclusions (Cascio & Valenzi, 1977).

Rater Sex. Finally, research on rater sex has revealed that women may be slightly more lenient than men when assigning performance ratings. In evaluating leadership style, women tended to rate their superiors higher than men on dimensions such as "structuring behavior" (Bartol & Butterfield, 1976). Also, student ratings of paper people yielded a similar trend. Women were found to rate individuals higher than men, and were seen as more confident when rating groups of people (London & Poplawski, 1976). Other past researchers though have asserted there to be no consistent main effect of sex on performance ratings (Landy & Farr, 1980). Instead, they see the interaction of sex with contextual variables within the organization as the more significant driver in performance rating variation (Landy & Farr, 1980). Both claims have been supported in subsequent research, as numerous studies have been unable to replicate these rater sex difference effects (Peters et al., 1984; Pulakos & Wexley, 1983). Even more convincing support was found in a study conducted in 1988. Results initially indicated a significant main effect of sex on performance ratings, with mean ratings from females actually

being lower than males. This effect disappeared though in subsequent analysis when the influence of contextual variables such as rating experience was controlled for (Northcraft, Huber, & Neale, 1988).

Other Rater Characteristics

Aside from simple demographics, many other (and often more complex) characteristics of raters have been shown to impact performance ratings. For instance, Borman and Hallam (1991) performed a study on 79 United States Air Force jet engine mechanics. The authors examined an array of individual difference correlates of observation accuracy and criticalness across two videotaped tasks. They found the most accurate raters to be those high in cognitive and spatial ability. The critical raters tended to be more experienced and less flexible/open to change (Borman & Hallam, 1991). What follows here are more specific categories of rater differences and their impact on ratings.

Rater Mood. Several studies have indicated that rater mood has the potential to influence performance ratings. Sinclair (1988), in a laboratory study, found that raters experiencing positive moods gave more lenient ratings to ratees, whereas those experiencing negative moods were significantly more severe (Sinclair, 1988). Hypothesizing similar results, Fried, Levi, Ben-David, Tiegs, and Avital (2000) collected data from two samples. The first was a laboratory-based organizational simulation conducted in the United States, and the other consisted of 60 managers in an Israeli firm who actually rated subordinates as part of the organization's performance evaluation process. Like Sinclair's (1988) findings, negative mood was associated with lower performance ratings. Contrary to the prior study though, positive mood was not significantly related to higher evaluations (Fried et al., 2000).

Rater Attitude. Along with mood, rater attitudes have also evidenced a relationship with performance ratings. In particular, rater attitudes toward the organization and toward the performance appraisal system itself have been shown to play a role. Tziner (1999) showed that Continuance Commitment toward one's organization and a sense of trust in the appraisal process both affected ratings. Continuance Commitment was defined as a strong desire by the individual to be a part of the organization. This desire would remain active as long as the relationship was beneficial to individuals by providing them with incentives such as interesting work and a generous retirement package (Tziner, 1999). Tziner indicated that the higher raters scored on this scale, the more likely they were to use political considerations in the performance appraisal process. Since this type of commitment is characterized by egocentric motives (examples above), Tziner feels it is not surprising that these people would partake in such behavior as inflating ratings to obtain favors from others or to enhance the image of their workgroup. Those less likely to incorporate politics into their ratings trusted, believed in, and accepted the performance appraisal system utilized by their organization. Because they are less swayed by political considerations, these raters tended to provide more accurate ratings (Tziner, 1999).

Expanding on these findings, Tziner, Murphy, and Cleveland (2001) assessed the importance of both these kinds of attitudes in rating behavior. As hypothesized, they found that attitudes that are more proximal to the actual task of rating, such as those related to the appraisal process itself, have more of an impact than attitudes toward the organization, which are more distal (Tziner, Murphy, & Cleveland, 2001). The authors are quick to mention that just because these more distal attitudes do not correlate as strongly with rating behavior does not mean they are not important. However, if a goal of the organization is to influence rating behavior in some

way, these findings suggest that increasing raters' confidence in the appraisal system is the better place to start.

Cognitive Processing. Another aspect of the rater deals with the cognitive processes involved in evaluating subordinates. The majority of performance appraisal studies that would be labeled as cognitive are concerned with how research in human information processing can be used to draw valid generalizations about the evaluation of job performance (Murphy & Cleveland, 1995).

A good example of this is seen in Feldman (1981), who asserts that a supervisor must perform several cognitive tasks before actual performance appraisals are possible. First, they must recognize and attend to relevant information about employees. In this stage, if the behavior of an employee is consistent with the supervisor's expectations, it is noted and stored automatically. However, if a behavior departs from these expectations, the supervisor is then forced to pay conscious attention to it in order to analyze it further. Next, the information gathered in the first stage must be organized and stored for later access. Along with this, new information must also be integrated with previously obtained information. This is the stage in which prototypes and inferences are made about people. In other words, an employee is placed into categories and "further memory-based judgments of that employee are colored by the category prototype" (p. 130). As a result of this, under-evaluations or over-evaluations of employees can be made by associating the general evaluation of the category with the person. After this, relevant information must be recalled when judgments are required. This recall is thought to be biased based on how the information was initially categorized. Recall involves remembering the prototype of the relevant category, and features of this prototype are remembered as being true of the person regardless of whether these characteristics were actually

observed. Lastly, integration and judgment can take on two forms. One is cognitive integration, and deals with a rater cognitively piecing together bits of information that had been previously stored. The other, evaluative judgment, relates to affective states of the rater. For instance, if a supervisor experiences a series of negative events, this may make prior negative memories more accessible, leading to lower ratings (Feldman, 1981).

Applications of such cognitive models of performance appraisal can be seen in prior research dating before and after Feldman's article. For example, Schneier (1977) defined cognitive complexity as "the degree to which a person possesses the ability to perceive behavior in a multidimensional manner" (p. 541). He argued that this characteristic in raters would differentiate for whom behavioral expectation scales (BES) would be effective. BES are so thorough and comprehensive in their makeup that they place high cognitive demands on the rater (Schneier, 1977). As a result, it was hypothesized, and confirmed, that those high in cognitive complexity would produce ratings with less halo, leniency, and restriction of range when the cognitive demands of the rating format matched their level of cognitive complexity (Schneier, 1977).

Also, in looking at implicit personality theories, which are a type of schema or category mentioned by Feldman (1981), researchers found that performance judgments were impacted by inferred personality characteristics of the ratee. Even more convincing was the finding that these implicit personality theories played a role when the only available information to the rater was ratee behavior (Krzystofiak, Cardy, & Newman, 1988). In other words, raters placed individuals within these personality schemas and, when recalling information about them, made appraisals that were impacted by this categorization.

Rater Accountability/Anonymity. Another rater attribute that should be discussed deals not with an actual characteristic of the rater per se, but rather one the rater obtains via the performance appraisal system being employed. Certain aspects of appraisal systems provide varying levels of anonymity to raters, which has been shown to impact performance ratings (Antonioni, 1994). For example, when raters understand that they will have to explain the ratings they provided to ratees in a face-to-face feedback session, their ratings tend to be more positive (Klimoski & Inks, 1990). Explanations for these results usually center around rater accountability. By forcing the rater to provide such things as face-to-face feedback, they essentially have to "own up" to the ratings they give. As a result, it would seem logical that a rater who knows they have to explain themselves, and their ratings, would find it much easier to defend positive ratings. A rater who remains anonymous though would not have these same concerns, allowing them to provide lower ratings if they see fit. Past research in this area however has shown mixed results.

In a 1977 study, Stone, Rabinowitz, and Spool assessed the impact of rater anonymity in a university setting. Students were instructed to provide ratings of their professors, with anonymity being manipulated by having one group sign the cover sheet of the evaluation form. The authors were unable to find any significant differences in ratings between the anonymous and non-anonymous raters. The authors blame this unexpected finding on the fact that more than one professor was evaluated. They believed the professor to have a moderating effect, with some professors causing rater anonymity to lead to higher ratings, and some causing it to result in lower ratings (Stone, Rabinowitz, & Spool, 1977).

In an actual field setting, Antonioni (1994) evaluated the impact of rater anonymity and accountability on the upward appraisals made by subordinates in an insurance company.

Utilizing two procedures, one designed to make employees accountable for their ratings, and the other to have them remain anonymous, the authors found that subordinates assigned to the accountability scenario rated their managers much higher than those in the anonymous condition. Furthermore, subordinate opinion of the feedback process indicated that those who remained anonymous viewed the process more positively, a result that was supported by the higher subject attrition rates in the accountability scenario (Antonioni, 1994). An investigation of similar factors however, led Roch and McNall (2007) to find that increased anonymity actually led to higher ratings, which contradicts Antonioni's results.

Finally, the positive effects of rater anonymity/accountability were seen in an experimental setting. Mero and Motowidlo (1995) had participants view performance vignettes in which a subordinate (student actor) was made to perform well or poorly. These participants were made to feel accountable by having to justify their ratings of others to the experimenter in writing. This increased accountability led to more accurate ratings, suggesting that maybe it is who you have to justify the ratings to that matters. Translating this to an organizational setting, perhaps if a manager had to justify their ratings of subordinates to his or her boss, along with those being rated, increased accuracy would result.

Based on these prior studies, it is evident that rater characteristics, and their impact on performance ratings, have received much attention from researchers in the past. However, with so much attention being paid to rater characteristics in hopes of remedying rater errors, it seems an issue that has been overlooked is whether or not they are errors on the part of the raters at all. Although they do not deny the existence of these phenomena, some researchers (Murphy & Cleveland, 1995) argue that portraying them as mistakes may be misleading. Also, simply

assessing the impact of these individual characteristics on the so-called errors may be an approach that is too simplistic.

In an attempt to help address these very issues, Murphy and Cleveland (1995) developed a conceptual model of performance appraisal in organizations (see Figure 1 for a visual representation of the model). This model takes a different stance on what others have referred to as "errors," as will be seen in its description. The authors view appraisal as a communication and social process in which the rater is not a mere measurement instrument, but rather an agent actively pursuing specific goals (a view far removed from that described by Fletcher (2001)). The main assumption of their model is that appraisal outcomes (ratings) are a result of a rater's "goal-directed behavior" (p.18). Aiming for parsimony, the authors included only four components: the rating context, performance judgment, performance rating, and evaluation. What follows is an in depth description of the Murphy and Cleveland model and its components, along with an introduction to the present research effort, its goals, and hypotheses.

A Four-Component Model of Performance Appraisal

Component 1: Rating Context

Rating context refers to the organizational context in which ratings are collected. The dynamic nature of the rating process is denoted in this first component as it has the potential to impact the other three, along with how the rating data are eventually used. According to the authors, contextual variables can be separated into two categories, proximal and distal, with each offering different levels of influence on the rater.

The first category can be thought of as the environment within the organization itself, or the organizational environment. The variables within this level are very salient to the rater, and

have great potential to impact performance ratings because they operate in the proximal setting of the organization. The second level is best described as the organization's environment. That is, the actual external environment the organization is located in. Variables that operate at this level are more distal to the rater and include such things as the economic and cultural climates surrounding the organization (Murphy & Cleveland, 1995).

A prime example of a contextual variable (and one that will be discussed in greater detail later) is purpose of appraisal. Cleveland, Murphy, and Williams (1989) surveyed more than 100 organizations on their uses of performance appraisal. Of these, over 70% indicated that their main uses of performance appraisal came in making decisions regarding administrative issues (promotion, merit increases, etc.) and employee feedback and development. Although both of these purposes of appraisal are valuable, each may invoke different cognitive processes within raters, leading to different rating outcomes in each context (Murphy & Cleveland, 1995). These administrative and developmental rating contexts, and their influence on performance ratings, will be one of the main focuses of this research project.

Component 2: Performance Judgment

Performance judgment focuses on the judgments of raters who have observed or obtained information on the performance of their employees. It is important to note that judgments, although an intricate part of the rating process, are not the same as ratings. Judgments represent private evaluations of a ratee, whereas ratings are public statements about the same ratee's performance (Mohrman & Lawler, 1983, as cited in Murphy & Cleveland, 1995). According to this model of performance appraisal, judgments can be seen as partly context free, but also context bound in some ways. The context free aspects are those that lie within the individual

rater's cognition, regardless of the job-specific context the rater happens to be in. However, this job context does determine the specific judgments that must be made by the rater. The purpose of the appraisal (mentioned above), for instance, may influence what kinds of judgments are made. If the purpose of the appraisal is to provide an employee with feedback and offer developmental opportunities, the rater needs to tailor his or her judgments to provide information appropriate to that context (Murphy & Cleveland, 1995).

Component 3: Performance Rating

This model of performance appraisal differentiates between judgments and ratings because it views the rating of performance as a "process of goal-directed communication" in which "the rater uses performance appraisal as a tool to achieve well-defined goals" (Murphy & Cleveland, 1995, p. 26). For example, if a rater believes that the performance appraisal system is used mainly to decide who deserves a promotion, the rater may first decide who deserves the promotion and then fill out the appraisal form in a way that reflects this decision. This goal-directed behavior denotes that performance ratings may not involve any consideration of the ratee's actual present performance, and therefore may not be indicative of the rater's internal judgments of that performance (Murphy & Cleveland, 1995).

Under optimal conditions, the ratings provided will match up with the internal judgments of the raters. However, due to the fact that ratings are often used within organizations for numerous purposes (e.g. administrative and developmental), the rater may ignore judgments of actual performance and focus mainly on their rating goals associated with these settings.

Component 4: Evaluation

The final component, evaluation, refers to an actual evaluation of the ratings themselves. What do they mean? The issues of leniency and halo come into play in the evaluation of the ratings provided to ratees. These two concepts are traditionally viewed as rater errors that are detrimental to rating accuracy. However, under the pretenses of this performance appraisal model, leniency and halo are not considered errors at all, but rather consequences of rater goals. In other words, halo and leniency should not be considered mistakes on the part of raters, and accuracy may not be a valid criterion for evaluating performance ratings. Discrepancies between performance ratings and the actual performance of individuals do not represent an inability on the part of a rater to effectively rate performance. Instead, they portray goal-oriented decisions made by these raters, which are impacted by their present organizational context (Murphy & Cleveland, 1995).

It should be noted here that the integration of personality within this model involves three out of the four components proposed by the authors. Again, it is predicted that rater personality will interact with the rating context (component one) to influence performance rating (component three). This, in turn, should influence how people evaluate (component four) the performance ratings assigned to individuals in terms of what they actually mean (e.g., should something like rating inaccuracy be viewed as a mistake?). According to the authors, this proposed model is one way to organize performance appraisal research efforts. Also, it may serve as a springboard for posing research questions that attempt to address performance appraisal problems experienced within organizations.

Empirical studies based on this idea of goal-directed behavior have indeed provided some support for the goal-rating relationship proposed in the model. Murphy, Cleveland, Skattebo, and Kinney (2004) investigated the impact of rater goals on ratings in an academic setting. The

authors believed this type of environment would yield a conservative test of this goal-related hypothesis. Due to the fact that teacher ratings are anonymous and have few immediate consequences for the rater (students), if a significant effect was found in this setting it would offer an argument for even stronger links being present in scenarios that offer raters more incentives to give certain ratings (Murphy et al., 2004). Their results indicated that raters who possess different goals tend to give different ratings, even when observing the same performance, as each class evaluated the same professor. In addition, this study was also able to assess changes in rater goals by surveying the students at the end of the semester as well as at the beginning. Interestingly, the authors found that goals can change over time, as a function of the performance being evaluated. Specifically, it was noted that raters who evaluated their professor more favorably placed more emphasis on the goal of "conveying information about strengths" at the end of the semester than they did at the beginning. This suggests that raters may alter/adapt their rating strategies (attend to different goals) as they learn more about the actual performance of the ratee (Murphy et al., 2004).

Taking this a step further, Wong and Kwong (2007) evaluated the impact of different goal types specifically on mean ratings and discriminability of ratings within groups of students completing a group project. The authors established significant relationships between harmony goals and fairness goals. Harmony goals involved raters maintaining group harmony and interpersonal relationships. Fairness goals involved giving performance ratings that fairly and accurately depicted each individual's contribution to their groups. Raters who pursued both of these types of goals increased their mean ratings (were more lenient) and decreased their discriminability (Wong & Kwong, 2007). The authors offer a practical implication of these findings by asserting that specifying rater goals in a performance evaluation may be a useful

method to help minimize discrepancies among raters. Rather than inferring their own goals, raters would be told what their goals should be in a given scenario. This type of instruction seems like an attempt to calibrate raters and, in turn, their ratings as well.

With this in mind, it should be said that although much research exists that investigates various rater characteristics, there have been relatively few studies of the specific role that rater personality plays in performance appraisal (Tziner, Murphy, and Cleveland 2002), especially as it relates to this model and the notion of rating as goal-directed behavior. Also, as indicated by the model, "in order to understand more fully and predict the perceptions and behaviors of individuals, psychologists should systematically investigate the relevant features of the context within which people behave" (Murphy & Cleveland, 1995, p. 58). In other words, simply assessing the impact of a rater characteristic such as personality may not be adequate. Performance appraisal is not a process that takes place in a vacuum. In order to fully comprehend the actions of the rater, one must examine characteristics of the raters themselves, as well as those of the environment in which the raters act. What follows is a discussion of the scant research dealing with the impact of rater personality on performance ratings, along with a description of the theory that illustrates its connection with the Murphy and Cleveland model.

Rater Personality

Similar to past research on other rater characteristics, personality-based research in the field of performance appraisal has focused on the impact that personality traits have on so-called rater errors. For example, a study by Tziner, Murphy, and Cleveland (2002) investigated the role of conscientiousness, but as a moderator between rating attitudes and rating behavior. Specifically, it was hypothesized that raters' attitudes and beliefs regarding performance

appraisal, such as confidence in and comfort with the system, would be related to their rating behavior. This relationship however, would be dictated by the raters' level of conscientiousness. In support of this belief, results showed raters who were low in conscientiousness were more likely to provide performance ratings that reflected their attitudes and beliefs about the performance appraisal process. According to the authors, this finding could be attributed to the idea that high conscientiousness individuals simply are not influenced by such contextual factors as are low conscientiousness raters (Tziner, Murphy, & Cleveland, 2002).

In a multi-source feedback setting, self-ratings of performance seem to be impacted by rater personality. Goffin and Anderson (2006) found that self-raters' levels of self-esteem and achievement were associated with inflated ratings when compared to those given by their supervisors and peers. Self-raters' anxiety levels were negatively associated with inflated ratings in relation to supervisor ratings. Raters high in anxiety rated themselves lower than their supervisors rated them. The findings related to self-esteem coincide with past research (Baird, 1977; Shrauger & Terbovic, 1976), and shows that positive self-evaluations (high self-esteem) lend themselves to positive performance evaluations in an organizational setting.

Self-monitoring is another stable dispositional characteristic (Snyder, 1974) that has been shown to impact performance ratings. High self-monitors are prone to analyzing their social environment and tailoring their behavior to fit the social context they find themselves in. Due to the idea that lenient ratings are likely to gain the approval of ratees (Murphy & Cleveland, 1995), Jawahar (2001) hypothesized that self-monitoring would be negatively related to rating accuracy. Results of the study indicated that rating accuracy does indeed decline with higher scores on self-monitoring (Jawahar, 2001).

In a study that illuminates the complexity of the performance appraisal process quite well, Yun, Donahue, Dudley, and McFarland (2005) investigated the relationship of rater personality, rating format, and social context with performance ratings. It was discovered that raters high on agreeableness tended to elevate ratings when they expected to have a face-to-face feedback meeting. In addition, rating format moderated this relationship between agreeableness and rating elevation, with high agreeableness raters providing less elevated ratings when using a behavioral checklist than with a graphic scale. Lastly, and possibly most interesting, was the finding that when the ratee was a poor performer, even those high in agreeableness did not rate them more positively than they deserved. Thus, poor performance by an employee would not allow even those high in agreeableness to justify elevating their ratings (Yun, Donahue, Dudley, & McFarland, 2005).

Finally, Bernardin, Cooke, and Villanova (2000), in an attempt to predict leniency bias, hypothesized that the five-factor model (FFM) traits (Digman, 1990) of conscientiousness and agreeableness would both be predictive of rating elevation, but in opposite directions. Because high conscientiousness individuals are seen as careful and more thorough, their ratings were hypothesized to be negatively related to rating level. Those high in agreeableness however, have a desire for social approval and would therefore want to avoid the possible conflict that comes with providing low ratings. Thus, agreeableness was hypothesized to be positively related to rating level. As expected, those scoring high on agreeableness gave significantly elevated ratings of performance, whereas those who scored high on conscientiousness rated others significantly lower (Bernardin, Cooke, & Villanova, 2000).

In evaluating prior research regarding the impact of rater personality on performance ratings, two issues arise. First, with the exception of the study conducted by Bernardin and

colleagues, there is very little explicit mention as to how rater personality may tie into the goal-directed behavior talked about by Murphy and Cleveland (1995). Also, it seems that its focus has been primarily on FFM traits. Unfortunately, this may be problematic as other studies have portrayed the FFM, or Big 5, as an inadequate taxonomy of personality variables for predicting important criteria (Hough, 1992; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). In addition, authors fault the FFM for being only a description of personality that offers no theoretical basis for explaining the behaviors related to its dimensions (Hogan & Shelton, 1998). This study hopes to address these issues by examining the relationship between personality and rating behavior through the lenses of the Murphy and Cleveland (1995) model and the Socioanalytic Theory of personality (Hogan & Shelton, 1998). This theory ties in nicely with the model in that it is a goal-based/evolutionary view of personality, with personality traits motivating individuals' behaviors.

The Socioanalytic Theory of Personality

The Socioanalytic Theory of personality (Hogan & Shelton, 1998) has its roots in evolutionary psychology, in that it is based on two conclusions about human motivation that manifest themselves during social interactions, including those that occur in the workplace. The first is that humans desire to gain popularity. The other, and often conflicting goal, is to obtain status (Hogan, 1982). More recent publications refer to these motivations as attempts to "get along" and "get ahead" (Hogan, Rybicki, Motowidlo, & Borman, 1998). People get along by conforming to others, whereas getting ahead tends to lead to competition, with outperformance sometimes ending with jealousy and resentment (Hogan & Shelton, 1998).

The individual differences that most people commonly associate with varying levels of the Big 5 traits take shape in two different ways according to the Socioanalytic Theory. First, "people differ in terms of their temperaments," which are "heritable traits that give a characteristic emotional tone to a person's personality" (Hogan & Shelton, 1998, p. 131). Second, people tend to differ in the strategies they employ to get along and get ahead. Some of these behavioral strategies are more effective than others, with certain people being more likely to modify their less than effective ones (Hogan & Shelton, 1998).

These two forms of differences lead to two distinct definitions of personality. The first should be considered personality from the outside, and comes from the perspective of the observer. In other words, this is a person's *reputation*. The second is personality from the inside, or from the actor's perspective. This is referred to as *identity*, and equates to the person we think we are and that we hope others will believe we are (Hogan & Shelton, 1998). Hogan et al. (1998) suggest that our identities are what drive (motivate) our behaviors, which observers then perceive. Based on these behaviors, the actors are then labeled with descriptors such as "loyal," "dependable, " etc., and their reputations are formed.

Past research has utilized the Socioanalytic Theory to explain workplace behavior. For example, Conway (2000) identified several managerial performance development constructs and attempted to understand their motivational antecedents using personality correlates. One finding that came out of this study was that the Empathy dimension of the California Psychology Inventory (CPI; Gough, 1975, as cited in Conway, 2000) correlated with the performance construct called Interpersonal Effectiveness. Thus, according to Socioanalytic Theory, managers' high Empathy scores would indicate an empathetic identity. The presence of this

empathy would motivate them to develop strong interpersonal relationships at work so they may understand and share the feelings of others.

By exploring the tenets of the Socioanalytic Theory of personality, it is easy to see its connection with Murphy and Cleveland's model of performance appraisal. The model views raters as actively pursuing goals (goal-directed behavior) when rating the performance of other individuals. Similarly, the Socioanalytic Theory states that, depending on the characteristics that make up someone's identity, that person will be motivated to achieve certain goals so that his/her reputation coincides with his/her identity. This study hopes to strengthen this link between rater personality and rating behavior. As posited by Hogan, Hogan, and Roberts (1996), what people do is a function of what kind of people they are. In other words, people's behavior is a function of their personalities. Thus, rating behavior should not be an exception, and it is expected that a rater's personality will impact the ratings he or she provides,

A simple objective analysis of a manager's rating behavior can be easily obtained. Once data are collected, all one has to do to gain some insight into this behavior is examine the ratings a manager provides to his or her subordinates. Assessing a manager's personality in order to gain a deeper, more complete understanding of what motivates this behavior is not quite that simple. Personality measurement is any procedure that involves assigning numbers to a person's interpersonal style according to some set of rules (Hogan, Hogan, & Roberts, 1996). However, good personality measurement must possess two features. First, the measure should be reliable, with scores remaining consistent over time. Second, scores on the measure should relate to indexes of meaningful non-test behavior. Put simply, the personality score should predict real world performance (Hogan, Hogan, & Roberts, 1996).

The Hogan Personality Inventory

One such measure is the Hogan Personality Inventory (HPI), which will be utilized in this study (for scale development, reliability, and validity information see Hogan & Hogan, 1995). The HPI has its roots in the FFM, which is based on observers' descriptions of others. In other words, it concerns the reputations (mentioned above) of those being observed. The Socioanalytic Theory however, asserts that these reputations are a product of actions that are designed to establish, defend, or enhance a person's identity. Thus, in order to understand what might lead a person to a high or low standing on these five dimensions of reputation, one should ask what sorts of self-presentational behavior might serve as their antecedents. Although it has its origins in the FFM, the HPI was developed in line with the Socioanalytic Theory, and is therefore an assessment device that allows one to measure these self-presentational behaviors (Hogan & Hogan, 1995).

The items written for the HPI were intended to reflect the FFM model dimensions in this way. However, during scale construction, the authors noticed two things. First, the FFM dimension of Extraversion can be broken down into two, unrelated components. The first is Sociability, which reflects a person's impulsivity and need for social interaction. The second is Ambition, and concerns the competitive desire for status and leadership. The FFM dimension of Openness could also be separated into two distinct constructs, Inquisitive and Learning Approach, which reflect an individual's interest in culture and academic achievement respectively (Hogan & Hogan, 1995; see Figure 2 for a mapping of HPI subscales to FFM traits).

By tapping into the different personality dimensions put forth by the HPI, one should be able to gain an understanding of rater identity. These identities will influence their rating

behaviors, with the rater's goal being to match their reputations (as seen by others) to their internal identities (who they are individually). What follows is a description of the HPI scales and the corresponding hypotheses with regard to rating behavior. All information concerning the meaning behind each scale was obtained from the Hogan Personality Inventory Manual (Hogan & Hogan, 1995) and the Hogan Guide (Hogan, Hogan, & Warrenfeltz, 2007).

The HPI Scales

Adjustment. Adjustment refers to the degree to which a person is calm and self-accepting, or on the other hand, self-critical and tense. Lower scores on this scale reflect a tendency to be remorseful, unhappy, and stress prone. Higher scoring individuals are stable, calm, and handle pressure well. These types of people tend to be well-liked because their moods are so consistent. Since lower Adjustment managers tend to be prone to stress, they are more apt to provide higher ratings to their direct reports to avoid the stress that comes with assigning lower ratings.

> *Hypothesis 1a:* There will be a significant negative relationship between rater scores on Adjustment and ratings they assign to their subordinates.

Ambition. Ambition is the degree to which a person is socially self-confident, leader-like, competitive, and energetic. Lower scores here reflect someone who is willing to be a follower, and have tasks assigned to them. Higher scores are indicative of a person who is driven, competitive, and focused on success. This type of person may compete with their peers or subordinates to attain personal achievement. This competitive desire for personal achievement may cause managers to become critical of their direct reports, who they may take as a threat to their success.

Hypothesis 1b: There will be a significant negative relationship between rater scores on Ambition and ratings they provide to their subordinates.

Sociability. This scale measures the degree to which a person seems to need and/or enjoy interactions with others. Lower scores indicate being reserved, quiet, and shy, whereas higher scores equate to being gregarious, attention-seeking, and impulsive. Due to the value they place on social interactions, individuals who score higher on this scale should desire to avoid the conflict that may result from assigning lower performance ratings to their direct reports.

Hypothesis 1c: There will be a significant positive relationship between rater scores on Sociability and ratings they assign to their subordinates.

Interpersonal Sensitivity. This scale measures a person's perception, tact, and social sensitivity. Lower scoring individuals tend to be blunt, tough, and even insensitive. Those who score higher are diplomatic, friendly, and considerate, but tend to avoid confrontation. Similar to Sociability, the propensity to avoid conflict that characterizes higher scoring individuals on this scale should motivate them to avoid conflict whenever possible (e.g. conflict resulting from providing poor performance ratings).

Hypothesis 1d: There will be a significant positive relationship between rater scores on Interpersonal Sensitivity and ratings they assign to their subordinates.

Prudence. Prudence measures an individual's propensity to be conscientious, conforming, and dependable. Persons with lower scores seem unconventional, impulsive, and irresponsible. They also have a tendency to be impatient with details. Higher scoring people are reliable, thorough, conscientious, and attentive to detail. This kind of manager should exhibit this thoroughness when assigning performance ratings to his direct reports, resulting in more critical

ratings than someone who tends to be irresponsible and may not take the review process seriously.

Hypothesis 1e: There will be a significant negative relationship between rater scores on Prudence and ratings they assign to their subordinates.

Inquisitive. This scale deals with the extent to which a person is perceived as bright, creative, and concerned with intellectual matters. Lower scoring people are cautious of new ideas and experiences, and are thus uncomfortable in ambiguous situations. Those that score higher are creative, open-minded, and are problem solvers who think "outside the box." Due to the nature of this scale, and based on the tenets of Socioanalytic Theory, it seems that those who score higher or lower would not be motivated to rate a certain way. For example, just because someone is an open-minded problem solver (or the converse), it doesn't follow that this would make him/her rate their subordinates a certain way. This subscale may be suited to predict other managerial behaviors (i.e., problem solving ability), but performance rating does not seem to be one of them. Therefore, scores on this scale will not be included in the analyses.

Learning Approach. Learning Approach measures how much a person enjoys academic activities and values educational achievement for its own sake. Lower scoring individuals tend to view educational opportunities as something to be endured rather than enjoyed. Higher scorers value education and are usually up-to-date with current trends and developments within their profession. Since higher Learning Approach raters value educational opportunities, they may see performance ratings as a chance to provide their ratees with a learning experience. Therefore, their ratings may be more critical (lower) than those who score lower.

Hypothesis 1f: There will be a significant negative relationship between rater scores on Learning Approach and ratings they assign to their subordinates.

The main goal of this study is to expand our knowledge of the performance appraisal process. It is hoped that the hypothesized relationships above will shed light on how a specific rater characteristic can influence the ratings that are provided to others. It has been mentioned though that these ratings are not made in a vacuum. Organizations are dynamic environments, which possess characteristics of their own that have the potential to influence performance ratings.

As Murphy & Cleveland (1995) asserted, the rating context can be divided into two levels, each with differing degrees of proximity to the rater. As a reminder, the first level can be thought of as the environment within the organization itself. The variables within this level are very salient to the rater, and have great potential to impact performance ratings as they operate in the proximal setting of the organization. The second level is best described as the organization's environment. That is, the actual external environment the organization is located in. Variables that operate at this level are more distal to the rater and include such things as the economic and cultural climates surrounding the organization (Murphy & Cleveland, 1995). Although their importance in past research has been made clear (Baron & Sackett, 2008; Golden, Barnes-Farrell, & Mascharka, 2009; Korsgaard, Meglino, & Lester, 2004), it will be beyond the scope of this project to discuss in detail those factors found within the distal, external environment. Therefore, what follows is a description of the more salient organizational factors and their influence on performance ratings.

Rating Context

Organizational Politics. Politics within an organization often play a role in important organizational decisions and actions (Gioia & Longenecker, 1994). As it turns out, one of the actions that organizational politics has been shown to influence is the appraisal of performance. In a study conducted by Gioia and Longenecker (1994), 82 executives from eight manufacturing and service organizations were interviewed to gain their insight regarding the prevalence and effects of organizational politics. Thorough analysis of the responses yielded some disconcerting results. The authors summarized their findings into five main points. First, politics are prevalent in appraisal, and the higher you go in the organization, the more political the process becomes. Respondents indicated that as an executive rises, a greater number of factors, other than actual performance, affected their performance ratings. Second, because managerial work can be ambiguous and poorly defined, this fosters the development of the "executive rating game" (p. 51), in which superiors use their ratings to fulfill their own agendas. Third, performance is not the bottom line in executive appraisals. Rather, interviewees thought there to be three factors affecting the ratings they receive. These included the boss's agenda, the ratee's reputation, and the company's political climate. Fourth, senior executives have extreme latitude in evaluating subordinates' performance as they often have sole control over the appraisals. This allows them to give any ratings they see fit regardless of ratee performance, without the threat of backlash. Finally, the authors concluded that executive appraisal is a political tool used to control people and resources. Due to the fact that higher-level executives are fairly autonomous, they are in a position to influence performance ratings to protect their own self-interests (Gioia & Longenecker, 1994).

Much of what was found in these interview responses can be linked back to rater goals. Executives using the appraisal process to serve their own self-interests is a prime example of

how goals and rater motivations can manipulate performance ratings. It seems, at least in these cases, that "managers who use these instruments often actively pursue agendas that are incompatible with highly accurate ratings" (Gioia & Longenecker, 1994, p. 48).

Organizational Climate. One aspect touched upon in the previously discussed research deserves further mention. Some of the interviewees referenced the political climate of the organization as one of the factors, other than actual performance, that defined performance ratings. Although little research has been done investigating the influence of specific climates on performance appraisal (Murphy & Cleveland, 1995), it makes sense that these climates could potentially impact performance ratings.

According to Schneider (1985, 1990), climate is defined as the perceptions of the events, practices, procedures, as well as the kind of behaviors that get rewarded, supported, and expected. In other words, climate refers to what people expect from a certain setting or situation, based on what normally occurs at those times and places. Schneider, Salvaggio, and Subirats (2002) extended Schneider's idea into the organizational realm, saying that organizational climate is the average, or most common way that people within an organization feel about that organization. Given these definitions, it is not hard to relate this organizational climate variable back to performance ratings, through rater goals. If a manager knows that his or her organization places a great deal of emphasis on, and rewards employees for behaving a certain way, that manager will be motivated to assign high ratings on those dimensions of a performance review. In doing so, he or she makes his or her subordinates look good and, as was mentioned before, they in turn are cast in a positive light as well.

An example of this can be seen in a study conducted in 2008 by Lievens, Conway, and De Corte (2008). In assessing the relative importance of task, citizenship, and counterproductive performance on performance ratings, the authors revealed that in a team-based culture, raters assigned more weight to citizenship performance and less to task performance (Lievens, Conway, & De Corte, 2008). In explaining this finding, the authors assert that "raters might be capable of matching their rating policies to an organization's culture" (p.24). If raters know that the organization in which they work values team-based activities/behaviors, they will assign more weight to contextual performance in terms of its overall importance to job performance.

Multi-source context. It should be pointed out that, up to this point, when describing the rater, a strong emphasis has been placed on the role of an employee's manager. However, it is important to realize that other parties can serve as raters within this organizational context being discussed. In fact, past researchers admit that emphasis given to performance appraisal of subordinates by superiors overshadows the reality that other forms of evaluation frequently occur (Fox & Bizman, 1988). The driving force behind this reality is that multi-source and 360-degree feedback in organizational settings are experiencing high levels of popularity (van Hooft, van der Flier, & Minne, 2006).

Even as early as 1974, Borman suggested that raters at different organizational levels probably observe/attend to different facets of ratees' job performance. Therefore, the ratings from these different sources reflect these varying observations and possess low levels of agreement. Research since then has indeed shown that performance ratings can be impacted by this multi-source rating context.

In 2008, a study was conducted assessing the impact of self-other rating agreement on performance in a multinational population. Self-other agreement in this case was defined as agreement between the self-ratings of a focal manager, and ratings provided to the same focal manager by his or her peers and subordinates. Results showed more agreement between the others' (peer and subordinate) ratings of the manager than the managers' self-ratings and the ratings of these others. Interestingly, this study introduced cultural variables as moderators and found those in collectivist cultures to overestimate manager behaviors. Also, managers in low assertiveness cultures provided self-ratings that were more relevant to performance outcomes than those in high assertiveness cultures (Lewis, 2008).

Interested in ratings of counseling performance early on in training, Fuqua, Johnson, Newman, Anderson, and Gade (1984) assessed the comparability of manager, peer, and self-ratings. Their results indicated that peer ratings of performance were significantly higher than both self- and manager ratings, and the self- and manager ratings were not significantly correlated (Fuqua et al., 1984). Interestingly, the authors attribute the elevated peer ratings to possible social influences that may be involved in the rating process. More current researchers have attested to this by suggesting that peers are unwilling to differentiate between good and poor performers for fear of "rocking the boat" (Murphy & Cleveland, 1995). With regards to findings presented in the Fuqua and colleagues study, this rationale makes sense. All peers were counseling trainees and may have been hesitant to assign low ratings in fear of disturbing the peace amongst group members who were all trying to survive the same process.

Finally, in their discussion of this multi-source rating context, Murphy and Cleveland (1995) discuss and review literature relating to the implications these different rating sources have on performance ratings. In their overview, self, peers, managers, and subordinates are all

highlighted. However, one potential rating source that may have been overlooked is the outside evaluator, such as a consultant, who is a non-organizational team member with salient experience (Furnham & Stringfield, 1998). In a study that did include this viewpoint along with the traditional others, it was found that consultant ratings were nearly all the lowest and most critical (Furnham & Stringfield, 1998). Prior research has suggested that raters may inflate ratings if they feel uncomfortable partaking in such activities such as monitoring and appraising performance (Villanova, Bernardin, Dahmus, & Sims, 1993). Being outsiders, those such as consultants may have less emotional ties to an organization and its employees, thus allowing them to make these critical (and possibly more objective) ratings without as much hesitation.

Purpose of Appraisal. Although the significance of the three prior organizational factors cannot be denied, the most important contextual factor in understanding the appraisal process and its outcomes may be the purpose of the appraisal (Murphy & Cleveland, 1995). Performance appraisal can be used for many different things within organizations, and can vary from disciplinary action, to feedback/development, to administrative decisions such as promotion and retention (Zedeck & Cascio, 1982). If an appraisal system is used for one of these purposes, it may not yield the same outcome when the purpose is different (Murphy & Cleveland, 1995).

Past research in this area seems to support the above assertions. Taylor and Wherry (1951) conducted a study on military personnel in which they assessed the ratings of officers in an "experimental" condition and a "for keeps" condition (p. 39). In the experimental condition, raters were told the ratings they provided would be used for research purposes only. However, the ratings made in the "for keeps" condition possessed administrative implications. The authors' results showed that the average rating in this administrative condition was much higher than that in the experimental.

Zedeck and Cascio (1982) conducted a study in which they assessed the purpose of appraisal and rater training on the accuracy of ratings. Using policy-capturing techniques, they were able to evaluate how raters, who were assigned to randomized training and purpose groups, differ in their evaluations. Three hypotheses were tested. First, regardless of whether the raters were trained, the authors predicted three distinct clusters of raters would emerge, corresponding to the three purposes for which the appraisals were made (merit raise, development, and retention). Second, regardless of the purpose, they believed two distinct rater groups would emerge, one containing those who were trained, and the other being composed of those who were not. Finally, it was predicted that purpose of appraisal and training would both impact ratings, leading to six distinct clusters of raters representing the interaction of the three rating purposes and the two training groups. Results indicated that purpose of rating was the most important factor in explaining the variability in rating accuracy. Whether or not the raters were trained failed to show a significant main effect on accuracy. Further, the policy-capturing analyses revealed that rater strategy varies with the purpose of the rating. The same rating dimensions were weighted, combined and integrated differently depending on whether ratings were made for a merit raise, development, or retention (Zedeck & Cascio, 1982).

Interestingly, a similar study also investigated purpose of appraisal and rater training, and their impact on rating accuracy. The results however, told the exact opposite story from that of Zedeck and Cascio (1982). McIntyre, Smith, and Hasset (1984) employed a hiring purpose, a feedback purpose, and a research-only purpose combined with four levels of training. These included rater error training (training raters on common psychometric errors), frame of reference training (Bernardin & Buckley, 1981), a combination of the two, and no training at all. Accuracy in this case was assessed by comparing ratings made by study participants to "true scores"

derived from mean expert ratings. Although these authors found training to be the best predictor of rating accuracy, with frame of reference training significantly outperforming rater error training, their findings did support the idea that raters in the administrative decisions group would be less accurate (McIntyre, Smith, & Hassett, 1984).

In another study, Dobbins, Cardy, and Truxillo (1986) informed raters that their ratings would be used for either 1) scale validation, which defined an experimental purpose, 2) to provide the instructor with feedback, or 3) to make merit pay and promotional decisions. The authors were also interested in ratee sex, and assessed the impact of both purpose of appraisal and ratee sex on the accuracy of performance ratings. Their results indicated that male professors were rated as more effective than females, but only in the administrative decision purpose, and when they were rated by other males (Dobbins, Cardy, & Truxillo, 1986). This *purpose* x *sex* interaction not only sheds light on the importance of the purpose of the appraisal when performance is rated, but once again denotes the complexity of the process itself.

Past research in this realm of performance appraisal seems to lead to a common conclusion. That is, ratings made in administrative settings tend to be more lenient (higher) than those made for other purposes. This has been regarded as strong evidence that raters consciously distort their ratings when they believe these ratings to be tied to valued outcomes (Murphy & Cleveland, 1995). In the present study, the two purposes in question were developmental and administrative. The developmental ratings were meant for just that, employee career development. Unlike those made in the administrative setting, these ratings had no bearing on promotions, merit increases, etc. In other words, ratings assigned in administrative settings probably have more serious consequences than those in a developmental context. Poor ratings in the former could lead to an employee losing his or her bonus, and even result in termination.

The prior research presented indicates that it would be naïve to think that raters do not take this into consideration, and that it does not motivate them to assign ratings in certain ways in these contexts.

Personality (although not as measured by the HPI) has been linked to rating behavior, and purpose of appraisal, due to its known importance, has also been shown to impact performance ratings. However, no study to date has assessed the impact of both of these variables simultaneously. Do personality trends that exist in rating behavior remain constant in two different rating contexts, or should differences be expected? Do people who score higher in Ambition, for example, continue to rate their subordinates the same way when the ratings are being made for developmental reasons, as compared to administrative reasons? Since personality and purpose of appraisal have both been shown to impact rating behavior, purpose is expected to modify the relationships between aspects of personality and ratings that are given to subordinates.

This second prediction is merely an overarching one that rater personality and purpose of appraisal will interact in some way to influence rating behavior. However, the question that remains is how this interaction will take shape. It seems this could be investigated by examining each personality subscale within the HPI individually, and how the characteristics of the rating context may impact the behavior of those that score higher or lower on them. As was mentioned earlier, past research has shown that ratings made in an administrative setting tend to be higher. This is due (in part) to the fact that the consequences of these ratings on subordinates are much more severe than those of developmental ratings (Murphy & Cleveland, 1995). If a rater knows the ratings he or she is assigning have grave consequences for the careers of their direct reports, how might this interact with the personality characteristics of this same rater to influence their

rating behavior? What follows are hypotheses that unpack the general prediction made above. Each personality subscale, and how it might relate to purpose of appraisal, is broken down into more specific predictions based on the subscales' score interpretations and the severity characteristic of the rating context mentioned above. The multilevel nature of these relationships is then tested using SPSS Mixed Models analysis. This will be discussed in further detail later.

Adjustment. Lower Adjustment managers are much more prone to stress (Hogan, Hogan, & Warrenfeltz, 2007). Due to the gravity of their consequences, providing administrative ratings to subordinates will cause much more stress than making developmental ratings. Therefore, one could expect a much stronger negative relationship between Adjustment scores and ratings provided in the administrative setting than in the developmental setting. In other words, as Adjustment decreases, ratings in both contexts should increase, but a much stronger increase will be seen in the administrative setting.

> *Hypothesis 2a:* Rater Adjustment and rating context will interact such that there will be a stronger negative relationship between Adjustment and ratings made in the administrative setting, and a similar, but weaker relationship with ratings made in the developmental setting.

Ambition. Higher scorers on Ambition are very competitive and may even see their subordinates as a threat (Hogan, Hogan, & Warrenfeltz, 2007). If these subordinates were to receive higher ratings in the administrative setting, this could potentially lead to pay increases and a promotion, cementing their status as a legitimate threat to the manager. As a result, the administrative setting should produce a negative relationship between Ambition and ratings. Ratings made in the developmental setting though should not produce this same type of threat to managers.

Therefore, the ratings made in this context by higher and lower scorers on Ambition may not exhibit as much of a difference.

> *Hypothesis 2b:* Rater Ambition and rating context will interact such that there will be a negative relationship between Ambition and ratings made in the administrative setting, and a similar, but weaker relationship between Ambition and ratings made in the developmental setting.

Sociability. Higher scorers on this scale value social interactions and gregariousness (Hogan, Hogan, & Warrenfeltz, 2007), which may be put in jeopardy due to the severe consequences of giving lower ratings in the administrative setting. Therefore, higher Sociability raters will provide their subordinates with higher ratings in the administrative settings than those who score lower on Sociability. Although the developmental context should produce a similar relationship, the positive slope in this setting will not be as steep. Because of the value they place on social interaction, higher Sociability raters will still provide higher developmental ratings than those who score lower. However, since developmental ratings do not offer such grave consequences as administrative ones, the increase in rating level between higher and lower scorers will not be as strong.

> *Hypothesis 2c:* Rater Sociability and rating context will interact such that there will be a positive relationship between Sociability and ratings made in the administrative setting, and a similar, but weaker relationship between Sociability and ratings made in the developmental setting.

Interpersonal Sensitivity. Because lower scorers tend to be socially imperceptive and unconcerned with the morale of others (Hogan, Hogan, & Warrenfeltz, 2007), low scoring

individuals should consistently rate their subordinates lower. However, both rating contexts may produce a positive relationship between Interpersonal Sensitivity and performance ratings. Due to the severity of the consequences in the administrative setting though, the increase in administrative rating levels as Interpersonal Sensitivity increases should be greater.

> *Hypothesis 2d:* Rater Interpersonal Sensitivity and rating context will interact such that there will be a positive relationship between Interpersonal Sensitivity and ratings made in the administrative setting, and a similar, but weaker relationship between Interpersonal Sensitivity and ratings made in the developmental setting.

Prudence. Higher scorers tend to hold higher standards for their employees (Hogan, Hogan, & Warrenfeltz, 2007) so subordinates will have a hard time meeting these expectations and should receive lower ratings in both the administrative and developmental contexts than they would from lower Prudence managers. With the administrative ratings potentially yielding pay raises and promotions, higher Prudence raters may elevate these standards even more to ensure their subordinates are deserving of such rewards. Thus the negative relationship seen in the administrative context should be stronger than that of the developmental setting. In other words, there will be a negative relationship between Prudence and both administrative and developmental ratings, but the negative slope will be greater in the administrative context.

> *Hypothesis 2e:* Rater Prudence and rating context will interact such that there will be a negative relationship between Prudence and ratings made in the administrative setting, and a similar, but weaker relationship between Prudence and ratings made in the developmental setting.

Learning Approach. Lower scorers should be unconcerned with staff development (Hogan, Hogan, & Warrenfeltz, 2007) and would have less reason to assign lower ratings (lower ratings in either the administrative or developmental setting would indicate a need for development). Higher scorers would assign lower ratings in the developmental setting because this would present the best avenue for a teaching/learning experience. Thus, the negative relationship that exists between Learning Approach and ratings should be stronger in the developmental setting than the administrative setting.

> *Hypothesis 2f:* Rater Learning Approach and rating context will interact such that there will be a negative relationship between Learning Approach and ratings made in the administrative setting, and a similar, but stronger relationship between Learning Approach and ratings made in the developmental setting.

Method

Participants

The total sample consisted of 138 managers who rated 448 subordinates within a large, global consumer products company. Of these managers, 85% (n = 116) were male and 15% (n = 21) were female. Respondents ranged from 37 to 62 years of age with a mean age of around 49 years (*M = 49.03, SD = 5.54).* The same 138 managers' ratings were used in both the developmental and administrative settings, as each employee's manager rated them in both. On average, there were approximately three direct reports per manager in this data set.

Procedure

Data were collected as part of the organization's routine human resource processes. It should be mentioned that the developmental and performance feedback processes were indeed very distinct rating contexts. The company has different messaging attached to both to ensure managers understand that ratings made within each context will be used for very different reasons. Also, the administrative ratings provided during the performance feedback process occurred approximately five months after the developmental feedback ratings took place.

Developmental feedback ratings were made during the 360-degree feedback program. The organization uses this program for developmental purposes only, rather than for performance evaluation. The survey process was conducted online by an external survey consulting firm and respondents made ratings during normal work hours. As part of this process, managers were required to provide ratings to their direct reports on a 58-item leadership behavior survey. This resulting dyadic relationship within the multi-source setting will be discussed in greater detail later.

During this developmental process, rater personality data were also collected from each focal manager using the Hogan Personality Inventory (Hogan & Hogan, 1995). The organization partnered their developmental feedback process with personality assessment in order to gain a deeper understanding of managerial behavior. The developmental feedback allows managers to see "what" they do, while the Hogan Personality Inventory explains "why" they do it (based on the tenets of the Socioanalytic Theory of personality). Although this personality assessment is intended to shed light on why managers exhibit certain behaviors, within the realm of assigning performance ratings, it is this "why" that has not been previously investigated.

Administrative ratings were captured during the organization's performance management process, approximately five months after the 360-degree feedback process ended. This time delay (along with the messaging attached to each process mentioned earlier) resulted in establishing two, very separate, rating contexts. As part of this performance review process, each employee sets objectives for the year, which focus on behavior that directly impacts the success of the business itself. These objectives are subsequently approved by the manager and, at the end of the year, each manager meets with subordinates to discuss the extent to which these self-set objectives have been met. Administrative ratings were made by responding to a single-item, 5-point rating scale.

Measures

Rater Personality. To assess rater personality, the Hogan Personality Inventory (Hogan & Hogan, 1995) was utilized. This measure consists of 182 items divided across seven scales. These include Adjustment, Ambition, Sociability, Interpersonal Sensitivity, Prudence, Inquisitive, and Learning Approach, all of which were described above. Scores on these scales

are based on percentiles, with 65 to 100 being high, 36 to 64 being average, and 0 to 35 being low. Each scale is broken down into Homogenous Item Composites, or HICs (Zonderman, 1980, as cited in Hogan, Hogan, and Warrenfeltz, 2007). For a breakdown of the scales, HICs, and sample items please see Tables 2 through 8 below.

Developmental Ratings. These ratings were captured using 58 items developed specifically for use within this organization (See Appendix A for a list of the original 58-items). These items addressed various aspects of leadership effectiveness including, but not limited to, supporting and coaching others, forming strategies, and executing plans. Items were measured on a 5-point Likert scale ranging from 1 = *Small Extent* to 5 = *Great Extent*. Example items include "Looks ahead to reasonably anticipate business opportunities and obstacles," "Supports reasonable risk taking and allows for some failures along the way," and "Makes strategic decisions based on ethical considerations to ensure the long-term sustainability of the organization."

Administrative Ratings. Ratings were given on a 5-point rating scale, with 1 being the lowest possible rating and 5 being the highest. See Table 9 for a list of the scale scores and definitions. At the managerial level, performance ratings on these behaviors impact the annual bonus the manager receives.

Analysis and Results

The following section will first review descriptive statistics, including means, standard deviations, and correlations of demographic variables, HPI scales, leadership factors, and performance measures. Next, steps taken to assess the factor structure for the 58-item leadership survey will be presented. The one-with-many dyadic research design will be shown to offer the

appropriate avenue for investigating relationships between rater personality, rating context, and ratings of both developmental and administrative performance. Intraclass Correlation (ICC1) analysis will be reviewed, results of which provide justification for further inquiry into the proposed relationships. Finally, results of hypothesis testing using SPSS Mixed Models analysis to evaluate the one-with-many dyadic relationships will be presented.

Descriptive Statistics

Means standard deviations, and correlations among the demographic variables, HPI scales, leadership behavior factors, and performance measures can be seen in Table 1.

Factor Analysis

To determine the appropriate factor structure of the 58-item leadership survey, an exploratory factor analysis was conducted using principle axis extraction and direct oblimin rotation. This method is superior to principal components analysis, as the latter is merely a data reduction technique. Principal axis factoring, on the other hand, aims to reveal any latent variables that cause manifest variables to covary (Costello & Osborne, 2005). Also, in the social sciences, there is a general expectation that there will be some degree of correlation among the factors. Thus, the oblique rotation should have rendered a more accurate and reproducible solution (Costello & Osborne, 2005). Although there are several forms of oblique rotation to choose from, research has shown that there is no widely preferred method, and all tend to produce similar results (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Thus, the direct oblimin method seemed appropriate.

The exploratory factor analysis of the item pool yielded the three-factor structure as the most interpretable. These three factors accounted for 42.00% of the variance. The criteria for

retaining items was (a) items possessed loadings of .40 or greater and (b) items that failed to load on any factor were removed. In total, two items were removed, leaving 56 items spread across the three factors (see Appendix B for a list of items and factor loadings). These resultant factors were labeled *Business Acumen*, *Inclusiveness*, and *Integrity*. There was no need for reverse scoring on any of the 56 items.

Business Acumen contains 31 items that measure a manager's knowledge of the business and ability to achieve results. Example items include "Creates effective and sustainable long-term strategies," "Confronts and works to resolve tough issues," and "Actively seeks and takes advantage of ideas, best practices, and solutions developed elsewhere." This factor possessed a Chronbach alpha of .94. Scale scores for Business Acumen were created by taking the average of these 31 items. See Appendix C for the list of items included in this and the other two factors.

The Inclusiveness factor contained 16 items and measured a manager's willingness and ability to foster an inclusive, balanced culture within the workplace. Example items included in this factor are "Champions diversity of thought, style, and perspective," "Fosters an environment where people feel comfortable speaking with truth and candor," and "Demonstrates openness to and respect for others' opinions and points of view." This factor possessed a Chronbach alpha of .91. Scale scores for Inclusiveness were created by taking the average of these 16 items.

The third factor, Integrity, contained 9 items that measure a manager's responsibility and degree of ethical behavior. Sample items include "Makes strategic decisions based on ethical considerations to ensure the long-term sustainability of the organization," "Acts with integrity on a daily basis even when it is difficult to do so," and "Creates an ethical culture by encouraging

and rewarding acting with integrity." This factor had a Chronbach alpha of .90. Scale scores for Integrity were created by taking the average of these 9 items.

One-With-Many Design

A one-with-many dyadic research design was chosen to assess the impact of rater personality and purpose of appraisal on managers' ratings of subordinates. Although there are a number of commonly used dyadic designs in the behavioral sciences (Wittenborn, Dolbin-McNab, & Keiley, 2013), the manager-direct report relationship captured by the data in both the developmental and administrative rating contexts lent itself to a one-with-many design (Kenny, Kashy, & Cook, 2006). In this type of design, "a person is in multiple dyadic relationships, but each of the person's partners is in a relationship with only that one person" (pg. 263).

This type of one-with-many design is a family in itself, comprising three data structures in which the data come from the one, the many, or both (Marcus, Kashy, & Baldwin, 2009). The structure of interest in this study though was the 1PMT, or one perceiver, many targets design (see Figure 3 for a visual representation of this type of one-with-many design). This is the most common one-with-many variant and is characterized by a focal person providing data about each of his or her partners (Kenny, Kashy, & Cook, 2006). In other words, all of the data are generated by the "one" (Marcus, Kashy, and Baldwin, 2009).

One-with-many data are hierarchically structured because partners are tied to this focal person (Kenny, Kashy, & Cook, 2006). In other words, partners (or subordinates/ratees in the case of the present research) are nested within a focal person (or a manager/rater). The hierarchical structure consists of two levels where Level 1 is the partner, or ratee, and Level 2 is the focal person, or rater (Kenny, Kashy, & Cook, 2006). This multilevel approach allowed for

testing the impact of the Level 2 predictor (rater personality) and the Level 1 predictor (purpose of appraisal) on ratings made in both contexts, as well as the cross level interaction between both predictors on the outcome (Kenny, D.A., personal communication, November 10, 2010)

Intraclass Correlation (ICC(1)) Analysis

In order to justify conducting statistical analyses of the interaction between rater personality and purpose of appraisal and their impact on developmental and administrative ratings, one-way analysis of variance (ANOVAs) were conducted. Rater was the fixed factor and ratings (whether they be developmental or administrative) served as the dependent variable. These ANOVAs examined whether there was statistically significant between-rater variability in the ratings. Information from the analysis (specifically the mean squares between, mean squares within, and the average number of ratees per rater) was used to calculate intraclass correlation(1), ICC(1) values, which are estimates of the extent to which performance ratings are affected by the rater making them. Specifically, ICC(1) analysis provides an estimate of the proportion of between-rater variability in performance ratings (LeBreton & Senter, 2008). If an ICC(1) value is not significantly different from zero, the relationships of interest would be inconsequential because the ratings from Rater X would be too similar to the same types of ratings given to a ratee by Rater Y. Thus, ICC(1) will ideally be significantly different from zero because this generally indicates that who one is rated by does matter in terms of the ratings actually received (Bliese, 2000).

The one-way ANOVA for developmental ratings on the Business Acumen factor was statistically significant, F (124, 274)=4.31, p <.05, and ICC(1)=.51. The one-way ANOVA for developmental ratings of Inclusiveness was also statistically significant, F (124, 274)=3.12, p

<.05, and ICC(1)=.40. Similarly, the one-way ANOVA for developmental ratings on the Integrity factor yielded statistical significance, F (124, 274)=11.76, p <.05, and ICC(1)=.77. Lastly, the one-way ANOVA for administrative ratings was statistically significant, F (108, 213)=1.92, p <.05, and ICC(1)=.22.

The ICC(1) statistics provide insight into the proportion of variability in ratings that is attributable to the rater. Taking Business Acumen as an example, the ICC(1) of .51 means that 51% of the variability in ratings of Business Acumen is attributable to the rater who is making those ratings. Since all values were significantly different from zero, the manager a worker is rated by does indeed impact the ratings received in both contexts, thus providing justification for further inquiry into relationships between rater personality, purpose of appraisal and ratings.

Multilevel Analysis

In their 2009 article, Marcus, Kashy, and Baldwin assert that "data from individual psychotherapy research have a natural fit with multilevel modeling because clients are nested within their therapists, and so the therapist is the upper level unit, and the client is the lower level unit" (pg.540). Translated to an occupational setting, one can see then how the proposed relationships in this study lend themselves to the same type of analysis. The therapist, or the upper level unit, becomes the manager. The clients would then equate to the direct reports who are nested under the manager the same way their counterparts are under their therapists.

With this in mind, developmental and administrative ratings provided by the managers would be the lower level outcome variable. Purpose of appraisal or rating context would then be the within-subjects lower level predictor, and rater personality would be the between-subjects upper level predictor.

The relationship between the single predictor at Level 1 and a single predictor at Level 2 can be seen in the equations below. The Level 1 equation is:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + r_{ij}$$

where y_{ij} is the outcome value of y for observation *i* in group *j,* β_{0j} is the intercept of the regression equation for group *j*, x_{1ij} is the observed predictor (purpose of appraisal) of observation *i* nested within group *j*, β_{1j} is the regression slope of *y* on x_1 within group *j*, and r_{ij} is observation- and group-specific residual (Preacher, Curran, & Bauer, 2006). The Level 2 equations are:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + u_{1j}$$

where w_{1j} is the observed predictor (rater personality) for group *j*, γ_{00} and γ_{10} are the fixed intercepts, γ_{01} and γ_{11} are the fixed regression coefficients for w_{1j}, and u_{0j} and u_{1j} are the residual terms (Preacher, Curran, & Bauer, 2006). Finally, substituting the Level 2 equation into the Level 1 equation results in the reduced form equation such that

$$y_{ij} = (\gamma_{00} + \gamma_{10}x_{1ij} + \gamma_{01}w_{1j} + \gamma_{11}x_{1ij}\,w_{1j}) + (u_{0j} + u_{1j}x_{1ij} + r_{ij})$$

It can be seen that the regression of the Level 1 slope on the Level 2 covariate results in a cross-level interaction between x_{1ij} and w_{1j} (purpose of appraisal and rater personality) with regression coefficient γ_{11} (Preacher, Curran, & Bauer, 2006). See Figure 4 for a visual representation of the cross-level effects between the Level 1 predictor, purpose of appraisal, and the Level 2 predictor, rater personality.

Results of Hypothesis Testing

Rater Personality

Hypothesis 1 stated that rater personality would impact the ratings a manager gave his or her subordinates. This hypothesis was unpacked further, with predictions being made concerning each subscale of the Hogan Personality Inventory and how scores on each would relate to ratings given to direct reports (see Tables 10 through 27 for regression coefficients and significance levels regarding each HPI subscale and its relationship to ratings). In order to test each hypothesis nested within Hypothesis 1, an aggregate was taken for ratings of each developmental factor (Business Acumen, Inclusiveness, and Integrity) and administrative ratings. SPSS Mixed Models analysis then assessed the impact of rater personality scores on each of these three aggregates. In other words, three separate analyses were conducted. First, the analysis was conducted with the aggregate of Business Acumen and administrative ratings, then the aggregate of Inclusiveness and administrative ratings, and finally the aggregate of Integrity and administrative ratings. These same three analyses were conducted to assess the impact of each HPI scale score on ratings the managers provided to their subordinates.

Hypothesis 1a stated that there would be a significant negative relationship between rater scores on Adjustment and ratings they assign to their subordinates. As seen in Tables 10, 11, and 12, the main effect of Adjustment was not significant for any of the three Mixed Models analyses that were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). Results of the data analysis did not provide support for Hypothesis 1a.

Hypothesis 1b stated that a significant negative relationship would exist between rater scores on Ambition and ratings they assign to their direct reports. As seen in Tables 13, 14, and 15, the main effect of Ambition was not significant for any of the three Mixed Models analyses that were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). Results of the data analysis provided no support for Hypothesis 1b.

In terms of Sociability, Hypothesis 1c stated that there would be a significant positive relationship between rater scores on Sociability and ratings they assign to their subordinates. As seen in Tables 16, 17, and 18, the main effect of Sociability was not significant for any of the three Mixed Models analyses that were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). No support for this hypothesis was found.

According to Hypothesis 1d, a significant positive relationship would exist between rater scores on Interpersonal Sensitivity and ratings they assign to their subordinates. As seen in Tables 19, 20, and 21, the main effect of Interpersonal Sensitivity was not significant for any of the three Mixed Models analyses that were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). There was no support for Hypothesis 1d in the data.

Hypothesis 1e stated that a significant negative relationship would exist between rater Prudence scores and ratings they assign to their subordinates. As seen in Tables 22, 23, and 24, the main effect of Prudence was not significant for any of the three Mixed Models analyses that

were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). There was no support for Hypothesis 1e in the data.

Hypothesis 1f stated that a significant negative relationship would exist between rater scores on Learning Approach and ratings they assign to their subordinates. As seen in Tables 25, 26, and 27, the main effect of Learning Approach was not significant for any of the three Mixed Models analyses that were conducted (for Business Acumen/aggregate, Inclusiveness/aggregate, and Integrity/aggregate, respectively). No support for Hypothesis 1f was found in the data.

In summary, analysis of the data did not provide support for the relationships posed in Hypotheses 1a through 1f. No significant main effects of rater personality on assigned ratings were uncovered through examining these relationships via Mixed Models analyses. Potential reasons for this general lack of support will be addressed in the Discussion section of this study.

Rater Personality x Purpose of Appraisal

Hypothesis 2 stated that rater personality and purpose of appraisal (or rating context) would interact to impact rating behavior. Again, this was an overarching hypothesis that was broken down further by examining each personality subscale of the HPI in conjunction with the purpose for which the ratings were made, be they developmental or administrative. With there being three distinct factors of developmental ratings, three separate analyses were performed within each subscale of the HPI to capture the effects of the predictor variables on the outcome. For example, the first factor to manifest from the developmental rating scale was Business Acumen. The multilevel Mixed Models analysis in SPSS was designed to capture the predictors' impact on this factor first, with subsequent analyses being conducted on each of the two other factors, Inclusiveness and Integrity.

Hypothesis 2a stated that rater Adjustment and purpose of appraisal would interact such that there would be a negative relationship between Adjustment and ratings in the administrative setting, and a less severe, but still negative relationship between Adjustment and ratings in the developmental context. As Adjustment levels decreased, rating levels would increase. The rationale here was that since ratings in the administrative setting come with more severe consequences to the ratee (the potential for termination, loss of merit increase, etc.) lower Adjustment raters, who would be heavily impacted by the stress of dealing with these consequences, would simply provide higher ratings in the administrative settings to avoid the increased stress. Results are presented in Tables 10, 11 and 12.

The only significant interaction found when assessing the impact of rater Adjustment and purpose of appraisal on performance ratings was found when the developmental factor of Business Acumen was included (p = .047, see Table 10). However, although the relationship between rater Adjustment and rating level was negative in the administrative setting, the same relationship was not seen in the developmental setting. In fact, as rater Adjustment increased, ratings of Business Acumen did as well.

In order to plot this, and all other significant cross-level interactions in this study, Preacher, Curran, and Bauer's (2003) online tool was employed. Figure 5 graphically depicts the above relationship between rater Adjustment level and purpose of appraisal.

Hypothesis 2b posited that rater Ambition and rating context would interact such that there would be a negative relationship between Ambition and ratings in the administrative setting, and a less severe negative relationship across the rating levels for all three developmental factors. In this case though, as rater Ambition increased, a decline in rating level would be seen

in both contexts, with a greater drop off experienced in the administrative setting. The rationale here centered around the competitive nature of higher Ambition raters and the increased severity of administrative ratings. Higher scorers here may see their subordinates as a threat to their own career success. If these direct reports were to receive higher ratings in the administrative setting, this could lead to pay raises and potentially a promotion, thus increasing the threat level to the manager. Results are presented in Tables 13, 14 and 15.

Only one significant interaction was uncovered in the Mixed Models analysis. When assessing the impact of rater Ambition and purpose of appraisal on ratings, only when the developmental factor of Integrity was included did the two predictors interact significantly (p = .030). Interestingly, although the relationship in the administrative setting was in the anticipated direction, with higher Ambition raters providing lower ratings, ratings of Inclusiveness actually increased with Ambition. A graphical representation of this cross-level interaction can be seen in Figure 6.

Hypothesis 2c stated that rater Sociability and purpose of appraisal would interact such that there would be a positive relationship between Sociability and rating level in the developmental setting, and a similar yet even more drastic positive relationship in the administrative setting. As rater Sociability levels increase then, so too would the ratings in both contexts, with a more pronounced increase seen in the administrative setting. Since higher scorers on this scale place a strong value on social interaction and gregariousness, it makes sense that they would not want to put these things at risk by providing their direct reports with lower ratings. Plus, with the gravity of the ratings in the administrative setting being greater, this type of rater would be even more motivated to behave this way when rating for this purpose. Results are presented in Tables 16, 17 and 18.

Significant interactions were found between rater personality and purpose of appraisal across the three factors of the developmental setting when Sociability was entered as the Level 2 predictor (p = .014, .005, and .004 respectively when Business Acumen, Inclusiveness, and Integrity ratings were assessed). However, trends in the data point in some unanticipated directions. Across all three developmental factors, as rater Sociability levels increased, so too did the ratings. In the administrative context though, in which an even sharper increase was expected, ratings actually went down as Sociability went up. Graphical representations for these three interactions can be seen in Figures 7-9.

Hypothesis 2d posited something quite similar to that of 2c. It stated that in terms of Interpersonal Sensitivity, rater personality and purpose of appraisal would interact such that a positive relationship would be seen in both rating contexts, with a more pronounced increase seen in the administrative setting. The thought process behind this hypothesis followed that for Sociability above. Higher scorers on this scale tend to avoid conflict and be friendly, diplomatic, and considerate. The conflict that could arise from providing lower performance ratings, especially ones for administrative purposes, would be something these raters would try to avoid. Results are presented in Tables 19, 20 and 21. No support for this hypothesis was found.

It was also hypothesized (2e) that rater Prudence and rating context would interact such that there would be a negative relationship (as Prudence increased, ratings would decrease) between Prudence and administrative ratings, and a similar, but less drastic relationship seen with developmental ratings. It was believed that since the rewards for higher administrative ratings are greater, higher Prudence raters would hold their subordinates to an even greater standard in this setting, making it even more difficult to receive any kind of elevated ratings.

Results are presented in Tables 22, 23 and 24. No support was found for this particular hypothesis.

It was predicted in Hypothesis 2f that Learning Approach and purpose of appraisal would interact such that there would be a negative relationship between Learning Approach and rating levels in the administrative setting, and a similar, but more drastic relationship present across the developmental rating factors of Business Acumen, Inclusiveness, and Integrity. In other words, as rater Learning Approach levels increased, rating levels in both contexts were anticipated to decline. The decrease that would manifest in the developmental setting though was predicted to be more severe. The rationale here was that higher scorers on this subscale value teaching/learning experiences. Due to the nature of the developmental context, higher Learning Approach managers would see these ratings as the ideal avenue with which to provide their subordinates with an opportunity to grow (i.e., provide them with lower ratings). Thus, the negative relationship seen in this setting would be enhanced compared to the administrative context. Results are presented in Tables 25, 26 and 27.

The only significant interaction between Learning Approach and purpose of appraisal was seen when considering the developmental factor of Inclusiveness ($p = .049$). However, the expected sharper decline in the developmental ratings was not realized. In fact, as rater Learning Approach increased, ratings of Inclusiveness actually went up as well. Figure 10 depicts this significant cross-level interaction graphically.

In summary, only partial support was garnered for Hypotheses 2a through 2f. A significant interaction between rater Adjustment and purpose of appraisal was realized when developmental ratings on the Business Acumen factor were included. However, the anticipated

negative relationship was not present in this rating context. Rater Ambition and purpose of appraisal also significantly impacted ratings, but the anticipated relationship was seen only in the administrative setting. Partial support was obtained for Hypothesis 2c, as the anticipated relationship between rater Sociability, purpose of appraisal, and ratings of performance was seen across all the developmental factors, but not in the administrative setting. Finally, data analysis provided additional partial support for Hypothesis 2f. The anticipated relationship was realized between rater Learning Approach levels and administrative ratings, but the opposite occurred in ratings of Inclusiveness. For a summary of all hypotheses and results, see Table 28. Potential explanations for these relationships are provided in the following Discussion section.

Discussion

Much attention has been given to the measurement of job performance in past research, and for good reason. It is essential to gain the best possible understanding of something that plays such an important role in personnel decisions made in the workplace. Just as the process of performance appraisal itself has evolved over the years, so too have psychologists' efforts to explain and understand it.

What manifested from this decades-long flurry of research activity was a slew of mixed results. In particular, when examining the influence of rater characteristics such as race, age, education, sex, etc., the waters became very muddy. In an attempt to add clarity to the picture, Murphy and Cleveland (1995) developed a model of performance appraisal in which the main tenets surrounded/challenged the established views of rating "errors" and the raters themselves. Raters to them were not merely measurement instruments striving for accuracy. Rather, they were viewed as agents who actively pursue specific goals, which would be reflected in the

ratings they provided. The actual rating of performance then became a "goal-directed behavior" (p. 18). Could it be that raters are actually rating their subordinates based on their own motivations? If so, the accuracy criterion that was focused on so heavily would need to be re-evaluated. In other words, if ratings are based on something other than actual subordinate performance, then the definition of an "accurate" rating needs to be changed, or perhaps completely done away with. As was mentioned, past researchers have actually called for a moratorium on the use of such things as halo and leniency as measures of rating effectiveness (Fox, Bizman, & Garti, 2005).

In fact, more recent efforts to understand the performance appraisal process are also calling for this shift. Spence and Keeping (2011) "suggest that the research literature move even further away from its focus on accuracy and consider that accuracy may not be the goal of raters" (p. 91). One caveat the researchers mention though is that much of the existing research examining rating behavior as a motivated phenomenon does so by examining specific goals or motives of raters. What is lacking, according to the authors, is an established theoretical framework within which to investigate rating behavior (Spence & Keeping, 2011).

The current study augments existing research on performance appraisal/rating behavior by examining it through the lens of a motivational framework that Spence and Keeping (2011) called for. Specifically, the Socioanalytic Theory of personality (Hogan & Shelton, 1998) was employed to provide a theoretical backdrop as to why managers may be motivated to rate their subordinates as they do. This theory tied in nicely to Murphy and Cleveland's (1995) model of performance appraisal, which views performance rating as a goal-directed, or motivated, behavior. With the Socioanalytic Theory as a backdrop, how the goals of raters take shape could potentially be explained. The study hoped to add clarity to the muddy waters (mentioned above)

surrounding rater characteristics by showing the impact rater personality can have on the ratings managers assign. Finally, this study attempted to assess the combined impact of rater personality and purpose of appraisal on rating behavior, something that, to the author's knowledge, has not been investigated in the past. In short, this study investigated whether the goals or motivations of a rater were influenced by the raters' personality and the purpose for which he or she was making the ratings. (which would, in turn, impact the ratings they provide to their subordinates).

Interpretation of Results – Main Effects

Unfortunately, no support was found when testing hypotheses relating rater personality to assigned ratings. This finding speaks to the true complexity of the performance appraisal process. Again, Murphy and Cleveland (1995) describe the act of rating (Component 3 of their performance appraisal model) as a "process of goal-directed communication" in which "the rater uses performance appraisal as a tool to achieve well-defined goals" (p.26). Based on the results of this study, it seems these goals are influenced by more than just the personality of a given rater. Murphy and Cleveland are also quick to point out that the rating context (particularly purpose of appraisal) should probably be considered one of the most important factors in determining rater goals. Within this population of raters, it seems that personality may matter somewhat differently depending on the rating context, or purpose for which the ratings are made. The multiplicative effects of rater personality and purpose of appraisal produced interesting trends that will be discussed further in the following section.

Interpretation of Results – Interactions

Data analyses did uncover evidence for combined effects of rater personality and purpose of appraisal on performance ratings. Six significant cross-level interactions were observed,

although support for these combined effects can only be considered partial. When a hypothesized relationship was seen in one rating context, it failed to show up in the other. In fact, the evidence actually points in the opposite direction hypothesized.

An interesting trend arose concerning rater Sociability, the three developmental factors (Business Acumen, Inclusiveness, and Integrity), and administrative ratings. A significant interaction was found with all three developmental factors, but in each case a negative relationship was observed between Sociability and administrative ratings. These results seem counter-intuitive. Higher Sociability raters value relationships and gregariousness, and should therefore be more apt to provide higher ratings so as not to jeopardize these things. In other words, a positive relationship was expected. On top of this, providing lower administrative ratings would put them in even greater jeopardy than developmental ratings since consequences of ratings made in the administrative setting are grave. In order to preserve their relationships with subordinates, it would make sense for raters to provide even higher ratings in this setting than they are already expected to give in a developmental one.

A possible explanation for this counter-intuitive finding comes from the assertion that employee performance is dynamic and changes over time (Deadrick, Bennett, & Russell, 1997). When faced with providing an overall rating of performance, raters are charged with the task of integrating these changes in performance into one summary rating (Reb & Greguras, 2010). In a study conducted in 2010, Reb and Greguras hypothesized that ratee mean performance would impact administrative ratings more than developmental ratings. Their rationale was that since administrative ratings are often used to make compensation-related decisions, mean performance should be especially salient to raters (Reb & Greguras, 2010). Support for this hypothesis was indeed found in their study. Translating this to the current study, remember that the

administrative ratings were provided six months after the developmental ones. If a rater perceived fluctuations in performance in either direction during this time span, assigned administrative ratings could reflect this.

Relating this idea of dynamic performance to the Murphy and Cleveland (1995) model of performance appraisal, this counter-intuitive finding relating to Sociability could be the raters' performance judgments (Component 2 of their model) coming into play. Remember that performance judgment focuses on the judgments of raters who have observed or obtained information on the performance of their employees. These judgments are private evaluations, whereas the actual ratings are public statements about the same ratee's performance. In this study, assigned (public) ratings may actually match subjective rater judgments. Depending on which direction an employee's performance was trending, their manager could have assessed this information, and then integrated this judgment on the performance trend into a rating that reflects a summary of his or her judgment. This could have happened even with higher Sociability raters, who were expected to provide higher ratings due to their personality characteristics. Past research by Yun, Donahue, Dudley, and McFarland (2005) offers some support for this assertion. These authors found that when ratee performance was poor, even raters higher in Agreeableness did not rate them more positively than they deserved. The same could potentially hold true then for higher Sociability raters, explaining the lower administrative ratings in this study.

Reb and Greguras's (2010) take on dynamic performance may also shed light on the unanticipated results dealing with rater Learning Approach and performance ratings. A negative relationship was predicted between Learning Approach and ratings. For example, a higher Learning Approach rater would assign lower ratings since this would provide an avenue for employee improvement, something this type of rater would value. This relationship was

expected to be exaggerated in the developmental setting since the purpose of ratings in this setting deals directly with subordinate development/improvement. Contrary to this reasoning, the significant interaction found between rater Learning Approach, developmental ratings of Inclusiveness, and administrative ratings evidenced a positive relationship between Learning Approach and developmental ratings.

Looking at these results through the lens of dynamic performance, this positive relationship may not be that surprising after all. If, leading into the developmental feedback, a manager perceives (or makes a subjective judgment) that a subordinate's performance was trending upward, this may be evaluated positively as a sign of acknowledgement or to motivate further improvement. Conversely, deteriorating performance may be evaluated more harshly to send signals that a change is expected (Reb & Greguras, 2010). Once again, this potentially dynamic nature of subordinate performance may need to be taken into account when evaluating ratings. If an employee's performance is on the rise, someone who values learning and development (a higher Learning Approach rater) may want to see the trend continue and reward the employee with higher developmental ratings as a means to ensure it does.

Significant interactions were also found when investigating rater Adjustment and Ambition. In the case of Adjustment, personality and purpose of appraisal interacted when the developmental factor of Business Acumen was included. With Ambition, it was the developmental factor of Integrity that produced significance. In both cases, the hypothesized relationship was realized for administrative ratings. The relationship in the developmental setting though actually went in the opposite direction. It seems the impact of the more severe consequences presented to raters in the administrative setting could be showing through here. Lower Adjustment raters may have provided higher administrative ratings to alleviate the stress

they are prone to, and higher Ambition raters may have provided lower administrative ratings to help eliminate the threat of successful subordinates. Since these consequences are not as severe in the developmental setting, the anticipated rating patterns did not hold true.

In sum, it seems that studying the impact of rater personality alone may not be adequate when the criterion of interest is rating behavior. Rather, other factors, such as the purpose for which the appraisal is made, should also be investigated. Moreover, the unanticipated results of this study speak to a need for further research in both academic and applied settings in order to gain a firmer hold on the performance appraisal process, managerial rating behavior, and its determinants (this will be touched on in more detail below).

Limitations

As with most research efforts, there are several limitations of the current study. First, since this study investigated only one set of developmental and administrative ratings, its design is cross-sectional. Future research could examine ratings over several iterations of the performance appraisal process to uncover the longitudinal impact these two, and other factors have on ratings. In doing so, factors such as the dynamic performance of the subordinates mentioned above could also be included in the research.

One should remember that although each subscale of the HPI was investigated individually in this study, the managers here completed an entire personality profile. In other words, each manager had scores for all subscales within the HPI. The importance of this is illustrated in the last two significant interactions discussed in the previous section. Although the administrative ratings were in the hypothesized direction, the developmental ratings were not. Could a different picture be painted if a more holistic look at rater personality was taken? For

example, the same rater who is higher in Ambition may also be lower in Adjustment. What does the combined impact of just those two personality traits (and where the rater falls on their scales) look like? Future research could take this into account and assess which personality traits overpower, or trump others to impact rating behavior.

It should be noted that the tenuous relationships found in this study could be due to the highly restricted range of ratings in both ratings contexts. In terms of the developmental ratings, the vast majority were near the top of the rating scale (with respondents marking either a 4 or a 5). Remember that these ratings were collected as part of the organization's 360-degree feedback process. This type of range restriction is a known issue in just about all 360-degree feedback programs within organizations (Church, personal communication, April 10, 2015), and could be part of the reason for the lack of support in the proposed relationships.

With regard to the administrative context, ratings made for this purpose also evidenced some degree of range restriction as well. Although ratings made here tended to be lower (on average) than those in the developmental setting, it should be noted that very few direct reports received a rating of 5 or a rating of 1 (the top and bottom of the scale, respectively). Once again, this highly restricted range could have had an adverse effect on the ability to uncover significant relationships.

A potential remedy for this issue would be to collect performance measures that are not ratings-based, such as net sales or some sort of production output like cases shipped. These more objective measures would accomplish two things. First, they would remove some of the subjectivity from the rating process, allowing for more tangible ratings of performance. Second, it would allow an organization to tailor its administrative rating process to specific business units

based on the most relevant criteria for each unit. For instance, net sales could be utilized in a sales or marketing department, whereas cases shipped could be employed within a production group.

Implications

This study contributed to the literature on performance appraisal by shedding light on the impact of rater personality and purpose of appraisal on developmental and administrative ratings of performance, and doing so in an actual field setting. Data were gathered from working managers in a large US corporation rather than being collected in a laboratory setting with a student sample, which is typically done (Greguras, Robie, Schleicher, & Goff III, 2003).

Although only partial support was found for the hypothesized interactions, these results should not be ignored. For instance, it was found that raters who scored higher in Ambition did rate their subordinates lower in the administrative setting. As mentioned, this example could have serious implications. When evaluating subordinates' suitability for upward mobility, it is important for managers to keep in mind what their personality is characterized by, and how this may impact their rating behavior. A subordinate working under a higher Ambition manager could be slighted a job/promotion opportunity due to having lower ratings than another candidate. However, the question that would need to be asked is are these ratings related to the actual performance of the subordinate, or are they impacted more by the personalities of their managers and the purpose for which they are being rated? As Murphy and Cleveland (1995) are quick to point out, this then turns into a scenario where performance evaluations may not be based on the actual performance of a subordinate at all. Coaching from a certified Hogan coach may aid in such a situation as this could increase Managerial Self-Awareness (MSA), something

that has been linked to a heightened ability to assess one's own workplace behavior (Church, 1997).

Also, in light of some of the unanticipated results dealing with rater personality, purpose of appraisal, and their impact on ratings, it seems that an organization such as the one in question could benefit from taking a more holistic look at rater personality, as was mentioned in the Limitations section. This study's investigation into the impact of rater personality was very exploratory, as it was the first time Hogan Personality Inventory data were used in an attempt to predict rating behavior. Thus, a first step was taken to assess each subscale of the HPI individually, and their subsequent effects on rating behavior. I believe (and would suggest), that more personality data be collected at different managerial levels, and that a more holistic view of a manager's entire personality profile be studied before making firm conclusions on the HPI's ability to predict rating behavior in conjunction with other variables of interest.

In addition, given the vast complexity of the performance appraisal process, practitioners could utilize various aspects of their human resources processes to gain a better understanding of it. For instance, past research has shown that managerial attitudes towards the performance appraisal process itself can impact rating behavior (Tziner, Murphy, & Cleveland, 2001). If a company conducts survey research, such as an organizational health survey, it may want to consider adding items (if they are not present already) that allow for the assessment of these managerial attitudes. By partnering this applied survey research with personality and behavioral assessments (such as 360-degree feedback), a more complete explanatory picture of performance appraisal could be painted.

Directions for Future Research

To date, this author knows of no study that assesses the combined impact of rater personality and purpose of appraisal on performance ratings. It is hoped that this research will spark future inquiries into these two very important factors of the performance appraisal process.

Murphy and Cleveland (1995) describe the rating of performance as a "process of goal-directed communication" (p. 26). The Socioanalytic Theory of personality states that a person's identity influences their behavior so they may shape their reputation. Following from this, it would seem that rating behavior should be impacted. Based on the results of this study, it seems these goals are influenced, to an extent, by a rater's personality and the purpose for which they are making ratings. However, after this examination, the muddy waters are obviously still not completely clear. Further inquiry is warranted into what other factors may be involved in shaping the goals of raters, and thus influencing their rating behavior.

One such avenue may be to investigate the role gender plays in the relationship between rater personality, purpose of appraisal, and ratings. It has been shown that women tend to be more lenient than men in their appraisal of performance (Bartol & Butterfield, 1976; London & Poplawski, 1976). However, other researchers have shown gender to have mixed results when it comes to rating behavior (Landy & Farr, 1980; Northcraft, Huber, and Neale, 1988). How might this impact the relationships seen in this study? A more consistent (and more complete) picture may be painted if rater gender were investigated alongside these other variables also known to play a role in the performance appraisal process. Supplemental analysis was conducted on this population of managers to control for the effects of manager gender on the ratings they assign to subordinates. Although no significant results were obtained, it should be noted that the study population was comprised of 85% males and only 15% females. This lack of variability in

manager gender could have impacted the findings. For an overview of these results, see the collection of tables in Appendix D.

Similarly, the tenure of a manager could also be studied to assess how the length of time a person has spent with a company influences the ratings they provide their subordinates. Someone who is long tenured may have a better grasp on the culture within the organization. This insight into what gets rewarded/supported by the organization could lead to completely different rating behavior than that of a newly indoctrinated manager. If employee tenure data is not available, one could consider using age as a proxy as they are naturally dependent, time-related variables (Bedeian, Ferris, & Kacmar, 1992). However, according to their 1992 study, Bedeian, Ferris, and Kacmar found chronological age and tenure to be distinct variables, leading to different outcomes. In their research, tenure was uncovered as a more stable predictor of job satisfaction than age. Although it is not clear that the same would be true for rating behavior, this does seem to suggest that one should at least approach with caution when considering using age as a proxy for tenure.

Along with the demographic variables discussed above, the type of appraisal method utilized could also be examined. The organization involved in the current study employed the traditional appraisal method in which managers are asked to choose one scale point out of several Likert-type scale values (Fox, Bizman, & Garti, 2005). Fox, Bizman, and Garti (2005) agree with previous researchers (Deadrick, Bennett, & Russell (1997)) that fluctuation is a natural property of employee performance. Citing situational constraints and internal variation in energy, personal concerns, and health conditions as potential causes for this performance fluctuation, they also assert that the more traditional appraisal method just described may not be adequate to measure this variability (Fox, Bizman, & Garti, 2005). As an alternative, the authors

suggest the distributional assessment method (Kane, 2000, as cited in Fox, Bizman, & Garti, 2005), where raters are asked to record the estimated frequency of different levels of their subordinates' behavior. So, instead of choosing a single point on a Likert scale, a manager would assess the percentage of time their subordinate was performing at each level. These percentages would then add up to 100 percent.

The authors found this distributional assessment method to be more effective than the traditional method in measuring performance. It would be interesting to employ this method in a study such as the current one. This could potentially lead to uncovering more significant relationships relating to dynamic (or fluctuating) performance like the ones described above in the relationships between Sociability, purpose of appraisal and rating behavior.

One caution to this approach is that the distributional assessment method is much more time consuming and elaborate than the traditional method (Fox, Bizman, & Garti, 2005). As a result, it is not commonly used in organizational settings where schedules (outside of taking the time to evaluate subordinate performance) tend to be tight (Fox, Bizman, & Garti, 2005). If this method were to be employed, the ideal organizational setting would have to be obtained, or a simulation study in a laboratory may need to be conducted. The former would obviously be superior in terms of external validity, but organizational constraints may make it unfeasible.

Finally, future research could also investigate the impact of organizational culture (which one could view as a reflection of the organization's "personality") on the performance appraisal process. This study was conducted within a single organizational setting. According to the classic Lewinian model though, behavior is a function of both personality and a person's environment (Organ & Bateman (1986), as cited in Berr, Church, & Waclawski, 2000). Also,

Murphy and Cleveland attest that contextual variables such as organizational culture definitely play a role in the performance appraisal process (Murphy & Cleveland, 1995). Thus, future research that includes data from several organizations, in which differing cultures (or environments) may be present, could prove beneficial in attempting to understand the appraisal of performance.

References

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47,* 349-356.

Baird, L.S. (1977). Self and superior rating of performance: As related to self-esteem and satisfaction with supervision. *Academy of Management Journal, 20*, 291-300.

Barron, L.G. & Sackett, P.R. (2008). Asian variability in performance rating modesty and leniency bias. *Human Performance, 21*, 277-290.

Barrett, R.R. (1966). *Performance rating*. Chicago: Science Research Associates.

Bartol, K.M. & Butterfield, A. (1976). Sex effects in evaluating leaders. *Journal of Applied Psychology, 61*, 446-454.

Bedeian, A.G., Ferris, G.R., & Kacmar, K.M. (1992). Age, tenure, and job satisfaction: A tale of two perspectives. *Journal of Vocational Behavior, 40,* 33-48.

Bendig, A.W. (1955). Rater reliability and the heterogeneity of scale anchors. *Journal of Applied Psychology, 39*, 37-39.

Bernardin, H.J., Alvares, K.M., & Cranny, C.J. (1976). A recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology, 61*, 564-670.

Bernardin, H.J. & Buckley, M.R. (1981). Strategies in rater training. *Academy of Management Review, 6*, 205-212.

Bernardin, H.J., Cooke, D.K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85*, 232-234.

Bernardin, H.J. & Walter, C.S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology, 62*, 64-69.

Bliese, P.D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K.J. Klein & S.W.J.

Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco, CA: Jossey-Bass.

Borman, W.C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Performance, 20*, 238-252.

Borman, W.C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology, 60*, 556-560.

Borman, W.C. & Hallam, G.L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-difference correlates. *Journal of Applied Psychology, 76*, 11-18.

Butler, A.B. & Skattebo, A. (2004). What is acceptable for women may not be for men: The effect of family conflicts with work on job-performance ratings. *Journal of Occupational and Organizational Psychology, 77*, 553-564.

Cascio, W.F. & Valenzi, E.R. (1977). Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. *Journal of Applied Psychology, 62*, 278-282.

Chiang, F.F.T. & Birtch, T.A. (2010). Appraising performance across borders: An empirical examination of the purposes and practices of performance appraisal in a multi-country context. *Journal of Management Studies, 47,* 1365-1393.

Church, A.H. (1997). Managerial self-awareness in high performing individuals in organizations. *Journal of Applied Psychology, 82,* 281-292.

Cleveland, J.N. & Landy, F.J. (1981). The influence of rater and ratee age on two performance judgments. *Personnel Psychology, 34*, 19-29.

Cleveland, J.N., Murphy, K.R., & Williams, R.E. (1989). Multiple uses of

performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*, 130-135.

Conway, J.M. (2000). Managerial performance development constructs and personality correlates. *Human Performance*, 13, 23-46.

Cooper, W. (1981). Ubiquitous halo. *Psychological Bulletin, 90*, 218-244.

Costello, A.B. & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*, 1-9.

Crow, W.J. (1957). The effect of training upon accuracy and variability in interpersonal perception. *Journal of Abnormal and Social Psychology, 55*, 355-359.

Curtis, A.B., Harvey, R.D., & Ravden, D. (2005). Sources of political distortions in performance appraisals. *Group & Organization Management, 30,* 42-60.

Deadrick, D.L., Bennett, N., & Russell, C.J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management, 23,* 745-757.

DeNisi, A.S., Robbins, T., & Cafferty, T.P. (1989). Organization of information used for performance appraisals: Role of diary-keeping. *Journal of Applied Psychology, 74*, 124-129.

Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

Dobbins, G.H., Cardy, R.L., & Truxillo, D.M. (1986). Effects of ratee sex and purpose of appraisal on the accuracy of performance evaluations. *Basic and Applied Social Psychology, 7*, 225-241.

Ellis, A.P.J., Ilgen, D.R., & Hollenbeck, J.R. (2006). The effects of team leader race on performance evaluations: An attributional perspective. *Small Group Research, 37*, 295-332.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*, 127-148.

Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology, 74*, 473-487.

Fox, S. & Bizman, A. (1988). Differential dimensions employed in rating subordinates, peers, and superiors. *The Journal of Psychology, 122,* 373-382.

Fox, S., Bizman, A., & Garti, A. (2005). Is distributional appraisal more effective than the traditional performance appraisal method?. *European Journal of Psychological Assessment, 21,* 165-172.

Fried, Y., Levi, A.S., Ben-David, H.A., Tiegs, R.B., & Avital, N. (2000). Rater positive and negative mood predispositions as predictors of performance ratings of ratees in simulated and real organizational settings: Evidence from US and Israeli samples. *Journal of Occupational and Organizational Psychology, 73*, 373-378.

Fuqua, D.R., Johnson, A.W., Newman, J.L., Anderson, M.W., & Gade, E.M. (1984). Variability across sources of performance ratings. *Journal of Counseling Psychology, 31*, 249-252.

Furnham, A. & Stringfield, P. (1998). Congruence in job performance ratings: A study of 360 feedback examining self, manager, peers, and consultant ratings. *Human Relations, 51*, 517-530.

Gioia, D.A. & Longenecker, C.O. (1994). Delving into the dark side: The politics of executive appraisal. *Organizational Dynamics, 22*, 47-58.

Goffin, R.D. & Anderson, D.W. (2006). The self-rater's personality and self-other

disagreement in multi-source performance ratings: Is disagreement healthy? *Journal of Managerial Psychology, 22*, 271-289.

Golden, T.D., Barnes-Farrell, J.L., & Mascharka, P.B. (2009). Implications of virtual management for subordinate performance appraisals: A pair of simulation studies. *Journal of Applied Social Psychology, 39*, 1589-1608.

Gorman, C.A. & Rentsch, J.R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology, 94,* 1336-1344.

Gough, H.G. (1975). *Manual for the California Psychological Inventory* (rev. ed.). Palo Alto, CA: Consulting Psychologists Press.

Greguras, G.J., Robie, C., Schleicher, D.J., & Goff III, M. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology, 56,* 1-21.

Griffeth, R.W. & Bedeian, A.G. (1989). Employee performance evaluations: Effects of ratee age, rater age, and ratee gender. *Journal of Organizational Behavior, 10*, 83-90.

Hogan, J., Rybicki, S.L., Motowidlo, S.J., & Borman, W.C. (1998). Relations between contextual performance, personality, and occupational advancement. *Human Performance, 11*, 189-207.

Hogan, R. (1982). A socioanalytic theory of personality. *Nebraska Symposium on Motivation*, 55-89.

Hogan, R. & Hogan, J. (1995). *Hogan personality inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R., Hogan, J., & Roberts (1996). Personality measurement and employment decisions. *American Psychologist, 51*, 469-477.

Hogan, R., Hogan, J., & Warrenfeltz (2007). *The Hogan Guide*. Tulsa, OK: Hogan Assessment Systems.

Hogan, R. & Shelton, D. (1998). A socioanalytic perspective on job performance. Human Performance, 11, 129-144.

Hough, L.M. (1992). The "Big Five" personality variables – construct confusion: Description versus prediction. *Human Performance, 5*, 139-155.

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581-595.

Jawahar, I.M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology, 86*, 875-883.

Jefferson, A.L. (2010). Performance appraisal applied to leadership. *Educational Studies, 36*, 111-114.

Kane, J.S. (2000). Accuracy and its determinants in distributional assessment. *Human Performance, 13,* 47-84.

Kane, J.S., Bernardin, H.J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal, 38*, 1036-1051.

Kenny, D.A., Kashy, D.A., & Cook, W.L. (2006). *Dyadic Data Analysis.* New York, NY: Cambridge University Press.

Klimoski, R. & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes, 45*, 194-208.

Korsgaard, M.A., Meglino, B.M., & Lester, S.W. (2004). The effect of other orientation on self-supervisor rating agreement. *Journal of Organizational Behavior, 25*, 873-891.

Kraiger, K. & Ford, J.K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70*, 56-65.

Krzytofiak, F., Cardy, R., & Newman, J. (1988). Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior.

Journal of Applied Psychology, 73, 515-521.

Landy, F.J., Barnes, J., & Murphy, K. (1978). Correlates of perceived fairness and accuracy in performance appraisal. *Journal of Applied Psychology, 63*, 751-754.

Landy, F.J. & Farr, J.L. (1980). Performance Rating. *Psychological Bulletin, 87*, 72-107.

LeBreton, J.M. & Senter, J.L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815-852.

Lewis, W.R. (2008). A multilevel investigation of culture on the relationship between self-other agreement and job performance (Doctoral dissertation, University of Connecticut, 2008). *Dissertation Abstracts International, 69*, 6461.

Lievens, F., Conway, J.M., & De Corte, W. (2008). The relative importance of task, citizenship and counterproductive performance to job performance ratings: Do rater source and team-based culture matter?. *Journal of Occupational and Organizational Psychology, 81*, 11-27.

London, M. & Poplawski, J.R. (1976). Effects of information on stereotype development in performance appraisal and interview contexts. *Journal of Applied Psychology, 61*, 199-205.

Luthans, F. & Peterson, S.J. (2003). 360-degree feedback with systematic coaching: Empirical analysis suggests a winning combination. *Human Resource Management, 42*, 243-256.

Marcus, D.K., Kashy, D.A., & Baldwin, S.A. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counseling Psychology, 56,* 537-548.

McIntyre, R.M., Smith, D.E., & Hassett, C.E. (1984). Accuracy of performance ratings

as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69*, 147-156.

Mero, N.P. & Motowidlo, S.J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology, 80*, 517-524.

Mohrman, A.M. & Lawler, E.E. (1983). Motivation and performance appraisal behavior. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory.* Hillsdale, NJ: Lawrence Erlbaum.

Mount, M.K., Sytsma, M.R., Hazucha, J.F., & Holt, K.E. (1997). Rater-ratee race effects in developmental performance ratings of managers. *Personnel Psychology, 50*, 51-69.

Murphy, K.R. & Balzer, W.K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology, 71*, 39-44.

Murphy, K. & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-oriented perspectives*. Newbury Park, CA: Sage.

Murphy, K.R., Cleveland, J.N., Skattebo, A.L., & Kinney, T.B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89*, 158-164.

Northcraft, G.B., Huber, V., & Neale, M.A. (1988). Sex effects in performance-related judgments. *Human Performance, 1*, 161-175.

Organ, D.W. & Bateman, T. (1986). *Organizational behavior: An applied psychological approach* (3rd ed.). Plano, TX: Business Publications.

Peters, L.H., O'Connor, E.J., Weekley, J., Pooyan, A., Frank, B., & Erenkrantz, B. (1984). Sex bias and managerial evaluations: A replication and extension. *Journal of Applied Psychology, 69*, 349-352.

Preacher, K.J., Curran, P.J., & Bauer, D.J. (2003, September). *Simple intercepts, simple*

slopes, and regions of significance in HLM 2-way interactions. Retrieved from http://quantpsy.org/interact/hlm2.htm.

Preacher, K.J., Curran, P.J., & Bauer, D.J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31,* 437-448.

Pulakos, E.D. & Wexley, K.N. (1983). The relationship among perceptual similarity, sex, and performance ratings in manager-subordinate dyads. *Academy of Management Journal, 26*, 129-139.

Pulakos, E.D., White, L.A., Oppler, S.H., & Borman, W.C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770-780.

Raymark, P.H., Skowronski, J.J., Bevard, L.A., & Hamann, S.A. (2001). Influence of recorder affect on the content of behavioural diaries and the recall of behaviours. *Applied Cognitive Psychology, 15*, 373-393.

Reb, J. & Greguras, G.J., (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of Applied Psychology, 95,* 213-220.

Roch, S.G. & McNall, L.A. (2007). An investigation of factors influencing accountability and performance ratings. *The Journal of Psychology, 14*, 499-523.

Sackett, P.R. & DuBois, C.L.Z. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology, 76*, 873-877.

Schmidt, F.L. & Johnson, H.J. (1973). Effect of race on peer ratings in an industrial situation. *Journal of Applied Psychology, 57*, 237-241.

Schneider, B. (1985). Organizational behavior. *Annual Review of Psychology, 36*,

573-611.

Schneider, B. (1990). The climate for service: An application of the climate construct. In B. Schneider (Ed.), *Organizational Climate and Culture* (pp. 383-412). San Francisco: Jossey-Bass.

Schneider, B. Salvaggio, A.N., & Subirats, M. (2002). Climate strength: A new direction for climate research. *Journal of Applied Psychology, 87*, 220-229.

Schneier, C.E. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology, 62*, 541-548.

Sharon, A.T. & Bartlett, C.J. (1969). Effect of instructional conditions in producing leniency on tow types of rating scales. *Personnel Psychology, 22*, 251-263.

Shrauger, J.S. & Terbovic, M.L. (1976). Self-evaluation and assessments of performance by self and others. *Journal of Consulting and Clinical Psychology, 44*, 564-572.

Sinclair, R.C. (1988). Mood, categorization breadth, and performance appraisal: The effect of order of information acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes, 42*, 22-46.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*, 526-537.

Spence, J.R. & Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. *Human resource Management Review, 21,* 85-95.

Stone, E.F., Rabinowitz, S., & Spool, M.D. (1977). Effect of anonymity on student evaluations of faculty performance. *Journal of Educational Psychology, 69*, 274-280.

Taylor, E.K. & Wherry, R.J. (1951). A study of leniency in two rating systems.

Personnel Psychology, 4, 39-47.

Tziner, A. (1999). The relationship between distal and proximal factors and the use of political considerations in performance appraisal. *Journal of Business and Psychology, 14*, 1999.

Tziner, A., Murphy, K.R., & Cleveland, J.N. (2001). Relationships between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment, 9*, 226-239.

Tziner, A., Murphy, K.R., & Cleveland, J.N. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and rating behavior?. *International Journal of Selection and Assessment, 10*, 218-224.

van Hooft, E.A.J., van der Flier, H., & Minne, M.R. (2006). Construct validity of multi-source performance ratings: An examination of the relationship of self-, supervisor-, and peer-ratings with cognitive and personality measures. *International Journal of Selection and Assessment, 14*, 67-81.

Varma, A., Pichler, S. & Srinivas, E.S. (2005). The role interpersonal affect in performance appraisal: Evidence from two samples-the US and India. *International Journal of Human Resource Management, 16*, 2029-2044.

Villanova, P., Bernardin, H.J., Dahmus, S.A., & Sims, R.L. (1993). Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement, 53*, 789-799.

Wales, S. (2003). Why coaching?. Journal of Change Management, 3, 275-282.

Wittenborn, A.K., Dolbin-MacNab, M.L., & Keiley, M.K. (2013). Dyadic research in marriage and family therapy: Methodological considerations. *Journal of Marital and Family Therapy, 39,* 5-16.

Woehr, D.J. (1992). Performance dimension accessibility: Implications for rating accuracy. *Journal of Organizational Behavior, 13*, 357-367.

Woehr, D.J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology, 79*, 525-534.

Wong, K.F.E. & Kwong, J.Y.Y (2007). Effects of rater goals on rating on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology, 92*, 577-585.

Yun, G.J., Donahue, L.M., Dudley, N.M., & McFarland, L.A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment, 13*, 97-107.

Zedeck, S. & Cascio, W.F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. *Journal of Applied Psychology, 67*, 752-758.

Zonderman, A.B. (1980*). Inventory construction by the method of homogeneous item composites*. Unpublished manuscript, The Johns Hopkins University, Baltimore.

Table *1*. Means, standard deviations, and correlations for study variables.

		M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1.	Rater Age	49.03	5.54	-												
2.	Rater Gender	1.15	1.15	-.16	-											
3.	Adjustment	57.91	25.45	.04	-.07	-										
4.	Ambition	72.72	23.48	.11	-.09	.41**	-									
5.	Sociability	57.03	26.14	-.10	.03	.12	.40**	-								
6.	Interpersonal Sensitivity	49.06	31.90	-.05	.20*	.39**	.31**	.23**	-							
7.	Prudence	55.50	24.34	.04	.00	.34**	.17	-.17	.38**	-						
8.	Inquisitiveness	48.97	25.63	.03	.12	.14	.22*	.39**	.24**	-.06	-					
9.	Learning Approach	66.42	26.66	-.06	.06	.19*	.26**	.31**	.18*	.08	.24**	-				
10.	Business Acumen	4.07	.41	.15	-.11	.20*	.00	-.01	.06	.01	.02	.00	-			
11.	Inclusiveness	4.18	.48	.01	.11	.05	-.07	.15	.02	-.04	.04	.05	.50**	-		
12.	Integrity	4.54	.44	.11	-.04	.09	.03	.09	.01	-.05	.09	.08	.46**	.61**	-	
13.	Administrative	3.42	.76	.00	.09	-.16	-.04	-.01	.07	-.03	.08	-.07	.00	-.04	-.02	-

** $p < .01$ * $p < .05$

Table 2. HPI Adjustment scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Adjustment	Empathy	5	I would rather not criticize people, even when they need it
	Not Anxious	4	I am seldom tense or anxious
	No Guilt	6	I rarely feel guilty about some of the things I have done
	Calmness	4	I keep calm in a crisis
	Even-tempered	5	I rarely lose my temper
	No Complaints	5	I rarely complain about anything
	Trusting	3	People really care about one another
	Good Attachment	5	No matter what happened I felt my parents loved me

Table 3. HPI Ambition scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Ambition	Competitive	5	I am an ambitious person
	Self-confident	3	I am a very self-confident person
	Accomplishment	6	I am known as someone who gets things done
	Leadership	6	In a group I like to take charge of things
	Identity	3	I know what I want to be
	No Social Anxiety	6	I do not mind talking in front of a group of people

Table 4. HPI Sociability scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Sociability	Likes parties	5	I would go to a party every night if I could
	Likes Crowds	4	Being part of a large crowd is exciting
	Experience Seeking	6	I like a lot of variety in my life
	Exhibitionistic	5	I like to be the center of attention
	Entertaining	4	I am often the life of the party

Table 5. HPI Interpersonal Sensitivity scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Interpersonal Sensitivity	Easy To Live With	5	I work well with other people
	Sensitive	4	I always try to see the other person's point of view
	Caring	4	I am sensitive to other people's moods
	Likes People	6	I enjoy just being with other people
	No Hostility	3	I would rather not criticize people, even when they need it

Table 6. HPI Prudence scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Prudence	Moralistic	5	I always practice what I preach
	Mastery	4	I do my job as well as I possibly can
	Virtuous	5	I strive for perfection in everything I do
	Not Autonomous	3	Other people's opinions of me are important
	Not Spontaneous	4	I always know what I will do tomorrow
	Impulse Control	5	I rarely do things on impulse
	Avoids Trouble	5	When I was in school I rarely gave teachers any trouble

Table 7. HPI Inquisitive scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Inquisitive	Science Ability	5	I am interested in science
	Curiosity	3	I have taken things apart just to see how they work
	Thrill Seeking	5	I would like to be a race car driver
	Intellectual Games	3	I enjoy solving riddles
	Generates Ideas	5	I am a quick-witted person
	Culture	4	I like classical music

Table 8. HPI Learning Approach scale, HICs, and sample items.

Scale	HICs	# of Items	Sample Item
Learning Approach	Education	3	As a child, school was easy for me
	Math Ability	3	I can multiply large numbers quickly
	Good Memory	4	I have a large vocabulary
	Reading	4	I would rather read than watch TV

Table 9. Depicting the scale used to make ratings for administrative purposes within the organization.

Scale Score	Definition
1	Did not meet most key objectives: Unsatisfactory year overall; may have met some objectives, but did not meet expectations on the most important objectives.
2	Met most key objectives: Satisfactory overall; met expectations on most objectives.
3	Met key objectives: Good year overall; met key objectives and exceeded some objectives.
4	Exceeded key objectives: Very good year overall; exceeded expectations on the most important objectives and met expectations on others.
5	Significantly exceeded most key objectives: Outstanding year overall; significantly exceeded expectations on the most important objectives and met expectations on others.

Table 10. Regression coefficient and significance level for the main effect of rater Adjustment on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.722	.033	95.365	112.145	.000
Adjustment	.000	.001	96.688	.210	.834
Adjustment x Purpose	-.002	.001	64.186	-2.028	.047

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 11. Regression coefficient and significance level for the main effect of rater Adjustment on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.757	.032	92.755	116.416	.000
Adjustment	-.000	.001	94.024	-.747	.457
Adjustment x Purpose	-.001	.001	71.896	-.552	.583

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 12. Regression coefficient and significance level for the main effect of rater Adjustment on aggregate of administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.971	.032	93.948	123.296	.000
Adjustment	-.000	.001	95.881	-.373	.710
Adjustment x Purpose	-.002	.001	88.509	-1.855	.067

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 13. Regression coefficient and significance level for the main effect of rater Ambition on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.724	.033	95.147	112.151	.000
Ambition	-.001	.001	96.901	-.682	.497
Ambition x Purpose	-.002	.001	70.507	-1.796	.077

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 14. Regression coefficient and significance level for the main effect of rater Ambition on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.758	.032	93.111	116.708	.000
Ambition	-.001	.001	92.614	-.986	.327
Ambition x Purpose	-.001	.001	78.334	-.468	.641

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 15. Regression coefficient and significance level for the main effect of rater Ambition on aggregate of administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.973	.032	93.878	123.863	.000
Ambition	-.001	.001	96.661	-1.087	.280
Ambition x Purpose	-.003	.001	121.390	-2.191	.030

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 16. Regression coefficient and significance level for the main effect of rater Sociability on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.723	.033	95.548	112.028	.000
Sociability	-.002	.001	92.558	-1.284	.202
Sociability x Purpose	-.003	.001	58.123	-2.539	.014

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 17. Regression coefficient and significance level for the main effect of rater Sociability on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.756	.032	93.035	116.028	.000
Sociability	-.000	.001	89.648	-.247	.805
Sociability x Purpose	-.003	.001	59.775	-2.916	.005

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 18. Regression coefficient and significance level for the main effect of rater Sociability on aggregate of administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.970	.032	93.830	122.980	.000
Sociability	-.001	.001	91.548	-.796	.428
Sociability x Purpose	-.003	.001	83.637	-2.939	.004

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 19. Regression coefficient and significance level for the main effect of rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.723	.033	94.684	112.794	.000
Int. Sensitivity	.001	.001	100.339	1.119	.266
Int. Sensitivity x Purpose	.000	.001	62.383	-.068	.946

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 20. Regression coefficient and significance level for the main effect of rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.757	.032	92.548	116.504	.000
Int. Sensitivity	.001	.001	98.678	.538	.592
Int. Sensitivity x Purpose	.001	.001	71.563	.551	.583

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 21. Regression coefficient and significance level for the main effect of rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.971	.032	93.685	123.542	.000
Int. Sensitivity	.000	.001	98.293	.349	.728
Int. Sensitivity x Purpose	.001	.001	77.998	.505	.615

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 22. Regression coefficient and significance level for the main effect of rater Prudence on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.725	.033	96.436	111.500	.000
Prudence	.001	.001	88.308	.601	.549
Prudence x Purpose	.001	.001	52.605	.823	.414

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 23. Regression coefficient and significance level for the main effect of rater Prudence on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.758	.032	93.988	115.971	.000
Prudence	.000	.001	85.792	.230	.819
Prudence x Purpose	.001	.001	61.675	.516	.608

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 24. Regression coefficient and significance level for the main effect of rater Prudence on aggregate of administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.973	.032	95.165	122.912	.000
Prudence	.000	.001	87.635	.310	.757
Prudence x Purpose	.002	.001	71.676	1.240	.219

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 25. Regression coefficient and significance level for the main effect of rater Learning Approach on aggregate of administrative ratings and developmental ratings of Business Acumen.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.723	.033	95.344	111.903	.000
Learning App.	.001	.001	96.458	-.443	.659
Learning App. x Purpose	-.002	.001	56.027	-1.643	.106

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 26. Regression coefficient and significance level for the main effect of rater Learning Approach on aggregate of administrative ratings and developmental ratings of Inclusiveness.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.756	.032	92.188	116.034	.000
Learning App.	.000	.001	93.966	.085	.933
Learning App. x Purpose	-.002	.001	63.476	-2.003	.049

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 27. Regression coefficient and significance level for the main effect of rater Learning Approach on administrative ratings and developmental ratings of Integrity.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.971	.032	93.742	123.431	.000
Learning App.	.000	.001	94.241	.145	.885
Learning App. x Purpose	-.002	.001	74.082	-1.724	.089

Note. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 28. Summary of hypotheses and tests of significance.

Hypothesis	Results
1a: Main effect of Adjustment on ratings	*ns*
1b: Main effect of Ambition on ratings	*ns*
1c: Main effect of Sociability on ratings	*ns*
1d: Main effect of Interpersonal Sensitivity on ratings	*ns*
1e: Main effect of Prudence on ratings	*ns*
1f: Main effect of Learning Approach on ratings	*ns*
2a: Adjustment x Purpose interaction	*p*=.047 (Business Acumen)*
2b: Ambition x Purpose interaction	*p*=.030 (Integrity)
2c: Sociability x Purpose interaction	*p*=.014 (Business Acumen); *p*=.005 (Inclusiveness); *p*=.004 (Integrity)
2d: Interpersonal Sensitivity x Purpose interaction	*ns*
2e: Prudence x Purpose interaction	*ns*
2f: Learning Approach x Purpose interaction	*p*=.049 (Inclusiveness)

Note. * represents the aggregate of indicated developmental factor and administrative ratings

Figure 1: The four-component model of performance appraisal as depicted by Murphy and Cleveland (1995).

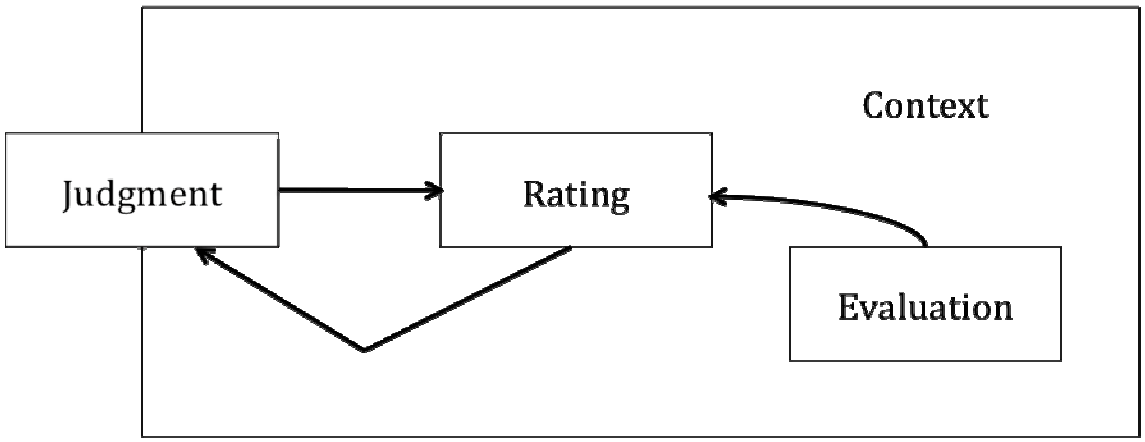
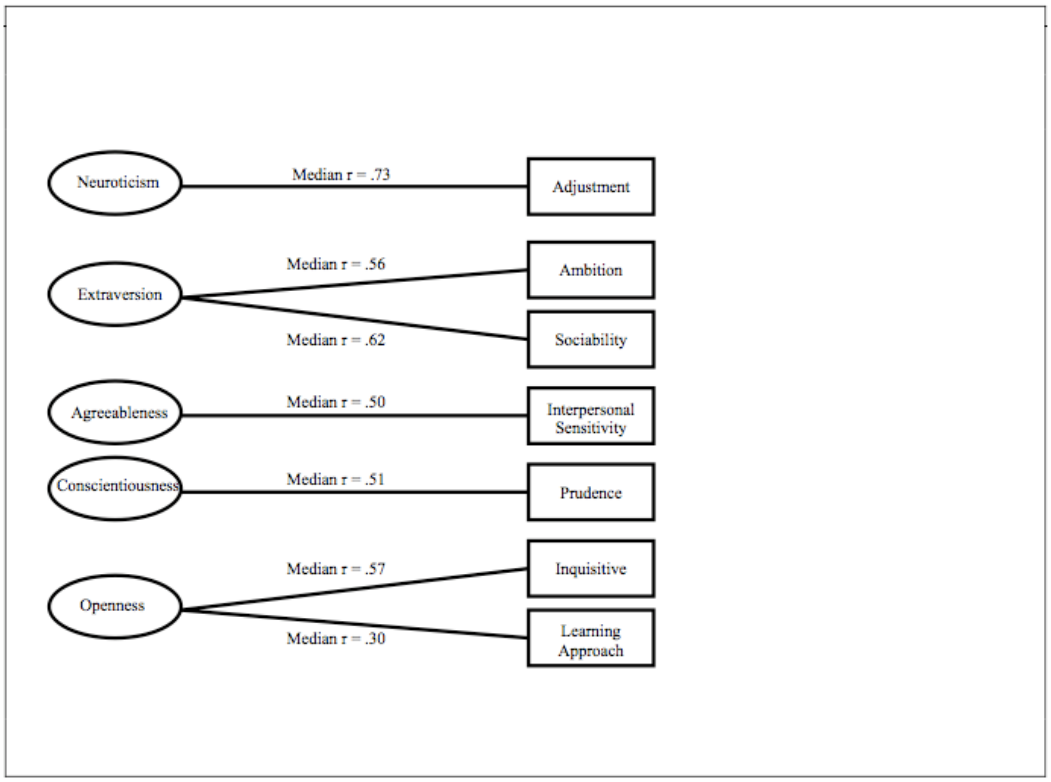


Figure 2: HPI mapping with FFM model traits and median correlations.



Note. Median correlation coefficients summarize HPI relations with the NEO PI-R (Goldberg, 2000), Goldberg's (1992) Big-Five Markers (R. Hogan & Hogan, 1995), Personal Characteristics Inventory (Mount & Barrick, 2001), and the Inventario de Personalidad de Cinco Factores (Salgado & Moscoso, 1999). The ranges of correlates are as follows: Adjustment/Emotional Stability/Neuroticism (.66 to .81), Ambition/Extraversion/Surgency (.39 to .60), Sociability/ Extraversion/Surgency (.44 to .64), Interpersonal Sensitivity/Agreeableness (.22 to .61), Prudence/Conscientiousness (.36 to .59), Inquisitive/Openness/Intellect (.33 to .69), Learning Approach/Openness/Intellect (.05 to .35).

Source: Benson (2006), Hogan and Holland (2003)

Figure 3: Illustration of the one-with-many design in which each focal person (raters A and B) is linked to partners (ratees 1 through 6), but these partners are not linked to any other focal person. More than one ratee is nested within each rater.

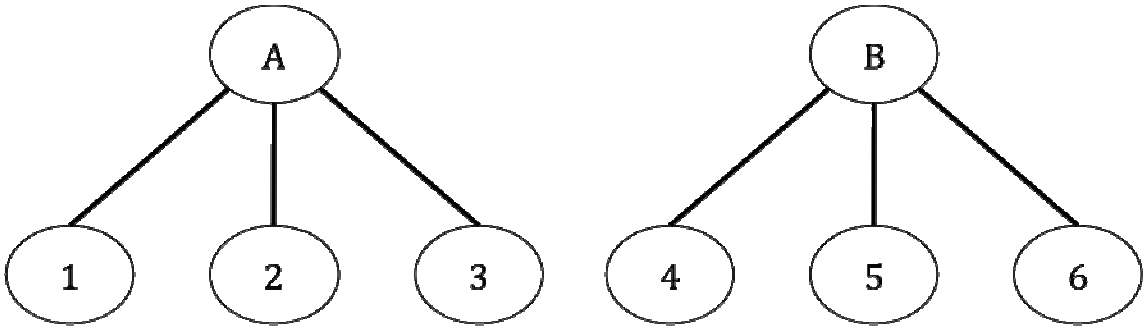


Figure 4: Visual representation of the cross-level effect of rater personality and purpose of appraisal on performance ratings.

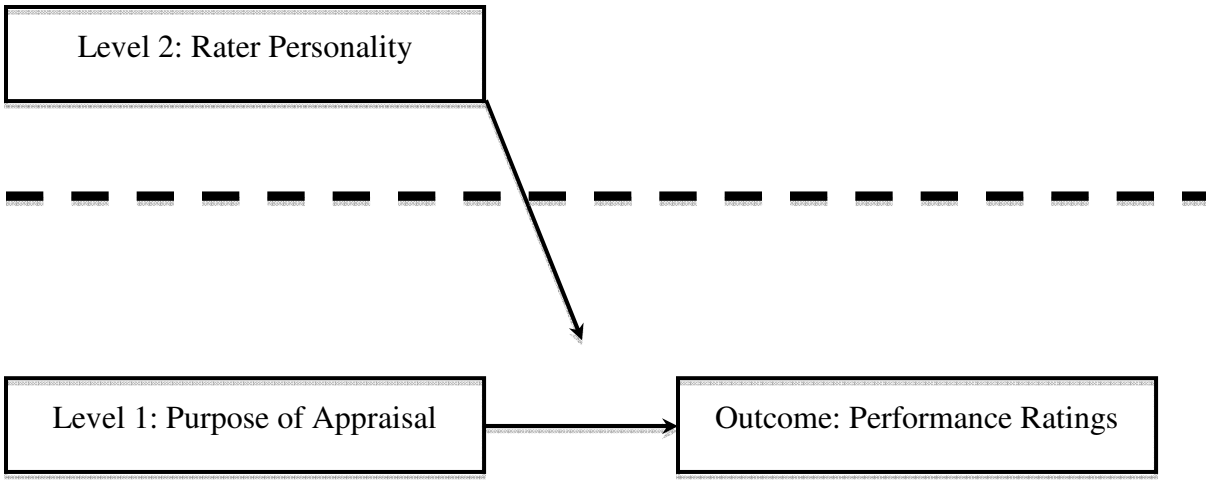
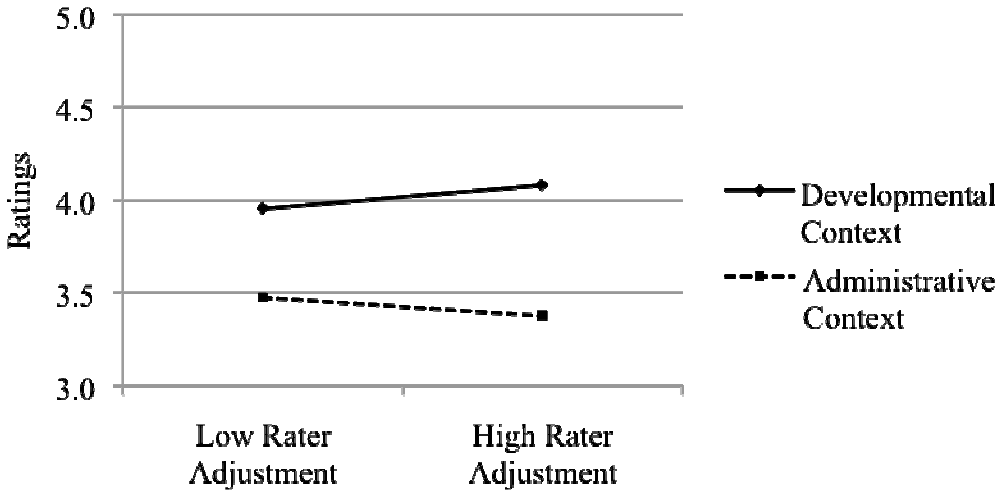
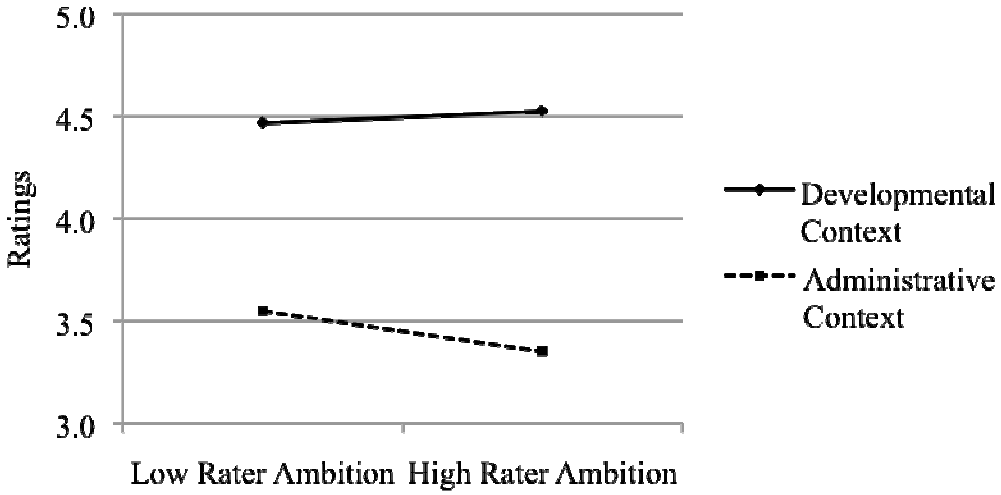


Figure 5: Graphical representation of the significant interaction between rater Adjustment and purpose of appraisal on administrative performance ratings and developmental ratings of Business Acumen.



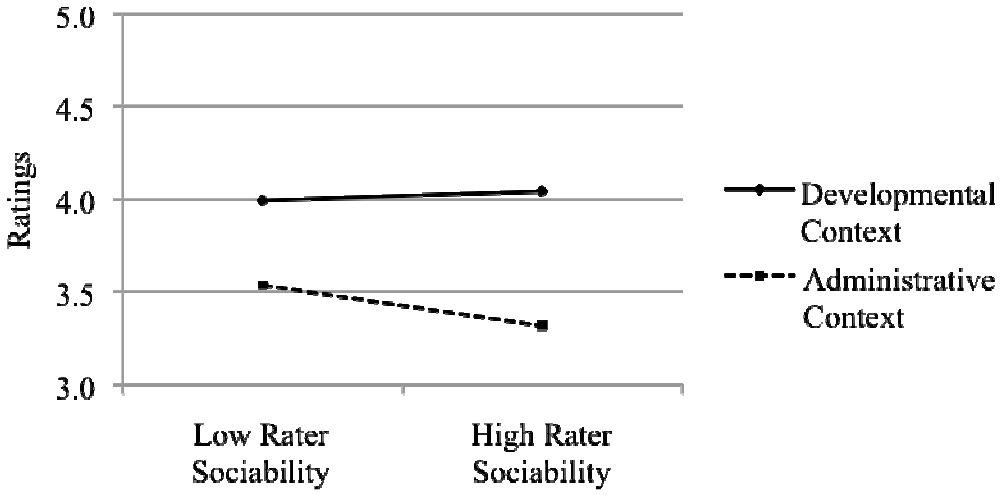
Note. Rater personality is not a categorical variable.

Figure 6: Graphical representation of the significant interaction between rater Ambition and purpose of appraisal on administrative performance ratings and developmental ratings of Integrity.



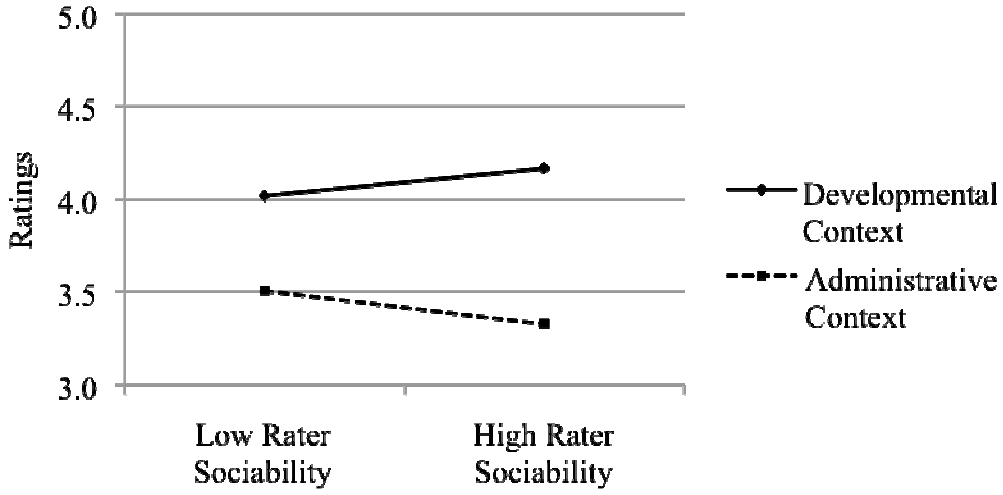
Note. Rater personality is not a categorical variable.

Figure 7: Graphical representation of the significant interaction between rater Sociability and purpose of appraisal on administrative performance ratings and developmental ratings of Business Acumen.



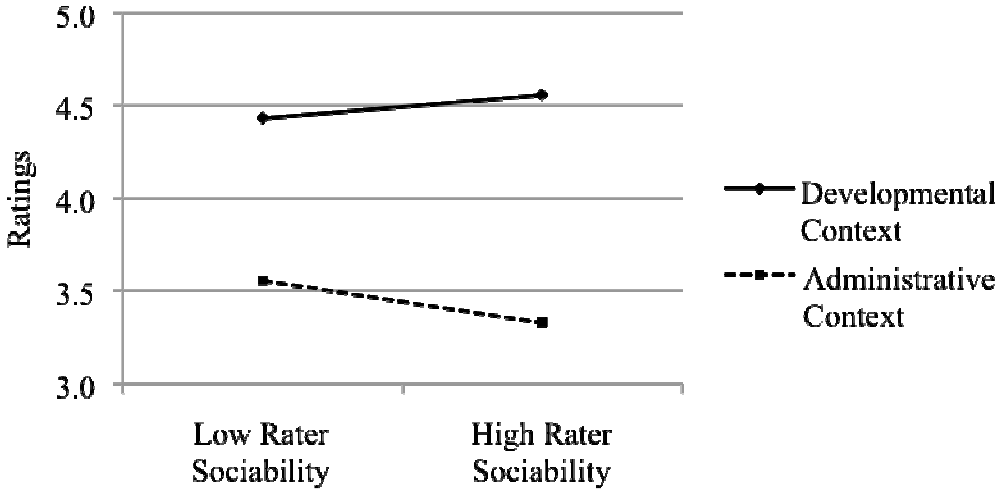
Note. Rater personality is not a categorical variable.

Figure 8: Graphical representation of the significant interaction between rater Sociability and purpose of appraisal on administrative performance ratings and developmental ratings of Inclusiveness.



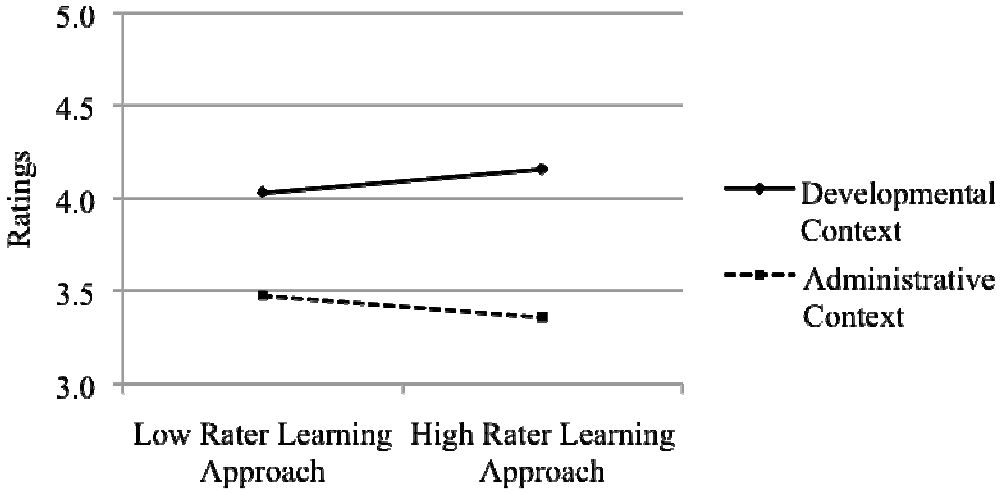
Note. Rater personality is not a categorical variable.

Figure 9: Graphical representation of the significant interaction between rater Sociability and purpose of appraisal on administrative performance ratings and developmental ratings of Integrity.



Note. Rater personality is not a categorical variable.

Figure 10: Graphical representation of the significant interaction between rater Learning Approach and purpose of appraisal on administrative performance ratings and developmental ratings of Inclusiveness.



Note. Rater personality is not a categorical variable.

Appendix A: Original 58-Item Leadership Questionnaire Used for Developmental Feedback Within Organization

1. Quickly analyzes complex problems to find actionable, pragmatic solutions
2. Sees connections in data, events, trends, etc.
3. Consistently works against the right priorities
4. Creates effective and sustainable long-term strategies
5. Demonstrates a broad perspective when solving problems and making decisions
6. Integrates and simplifies complex or competing information for timely decision making
7. Looks ahead to reasonably anticipate business opportunities and obstacles
8. Takes the initiative to find ways to get better results
9. Actively seeks and takes advantage of ideas, best practices, and solutions developed elsewhere
10. Develops creative solutions and different ways of doing things that add value
11. Effectively filters through and identifies the best ideas to pursue
12. Fosters an environment of innovation and personally champions breakthrough ideas and initiatives
13. Supports reasonable risk taking and allows for some failures along the way
14. Demonstrates perseverance and resilience in the pursuit of goals
15. Confronts and works to resolve tough issues
16. Exhibits a “can do” attitude and willingness to take on significant challenges
17. Makes the right trade-offs and balances resources to deliver results
18. Plans ahead and articulates the organizational vision and resource requirements needed to deliver results for the future
19. Demonstrates the courage and conviction needed to drive large scale change initiatives
20. Treats all people with respect and fairness

21. Demonstrates sensitivity to differences when dealing with people from different cultural backgrounds and/or other differences

22. Demonstrates openness to and respect for others' opinions and points of view

23. Champions diversity of thought, style, and perspective

24. Demonstrates sensitivity and awareness of cross-cultural implications when conducting business or executing initiatives

25. Creates a work environment that helps people achieve a healthy balance between work and personal life

26. Fosters a positive and inclusive work environment where all people feel respected and valued for their contributions

27. Demonstrates an awareness of his/her own strength and development areas

28. Seeks and acts on feedback from others regarding ways to improve his/her performance

29. Learns from experiences and can readily apply that knowledge to new situations

30. Seeks out learning opportunities and new experiences

31. Hires and develops people who have a positive impact

32. Demonstrates a commitment to the sharing of talent across divisions, functions, regions, and teams to build bench

33. Takes appropriate risks on people through the use of critical experiences and stretch assignments

34. Demonstrates a sense of passion, enjoyment, and pride about his/her work

35. Demonstrates a positive attitude in the workplace

36. Embraces and adapts well to change

37. Creates a work environment that makes work rewarding and enjoyable

38. Articulates a compelling vision, purpose, and direction that inspires others to follow

39. Demonstrates courage and optimism during times of crisis or change

40. Continually raises the bar for performance and helps others succeed

41. Collaborates well with others to deliver results

42. Keeps others informed so there are no unnecessary surprises

43. Effectively listens to and understands what other people are saying

44. Builds relationships outside his/her sphere of influence to get things done

45. Operates effectively in matrix relationships across organizational boundaries

46. Anticipates controversial issues and questions and provides effective solutions

47. Displays responsible and ethical behavior toward customers, suppliers, and others in the organization

48. Consistently follows the meaning and intent of company policies and standards (e.g., Code of Conduct)

49. Acts with integrity on a daily basis even when it is difficult to do so

50. Makes strategic decisions based on ethical considerations to ensure the long-term sustainability of the organization

51. Ensures systems, processes, and practices reinforce the PepsiCo values

52. Creates an ethical culture by encouraging and rewarding acting with integrity

53. Can be counted on to speak with truth and candor

54. Acts consistently with his/her words

55. Takes accountability for his/her own mistakes

56. Creates a safe and trusting work environment that encourages openness and honest dialogue

57. Builds trusting relationships with multiple internal and external stakeholders for the broader good of the organization

58. Fosters an environment where people feel comfortable speaking with truth and candor

Appendix B: Factor Loadings for Leadership Survey Items

Question	Factor		
	Business Acumen	Inclusiveness	Integrity
1	***.64***	--	--
2	***.59***	--	--
3	***.62***	--	--
4	***.67***	--	--
5	***.53***	--	--
6	***.64***	--	--
7	***.63***	--	--
8	***.65***	--	--
9	***.53***	-.40	--
10	***.62***	--	--
11	***.61***	--	--
12	***.51***	-.42	--
13	***.50***	-.44	--
14	***.52***	--	--
15	***.56***	--	--
16	***.56***	--	--
17	***.58***	--	--
18	***.68***	--	--
19	***.61***	--	--
20	--	***-.70***	-.44
21	--	***-.64***	--
22	--	***-.75***	--
23	--	***-.60***	--

24	--	***-.57***	--
25	--	***-.52***	--
26	--	***-.69***	--
27	***.52***	-.51	--
28	***.47***	-.45	--
29	***.61***	-.48	--
30	***.52***	-.44	--
31	--	--	--
32	--	--	--
33	***.43***	-.42	--
34	***.49***	--	--
35	--	***-.53***	--
36	***.54***	-.49	--
37	--	***-.63***	--
38	***.57***	--	--
39	***.52***	-.41	--
40	***.67***	--	-.46
41	.42	***-.66***	--
42	***.47***	-.45	--
43	.40	***-.63***	--
44	.43	***-.46***	--
45	.41	***-.52***	--
46	***.56***	--	--
47	--	--	***-.80***
48	--	--	***-.76***
49	--	--	***-.77***

50	--	--	***-.77***
51	.42	-.44	***-.79***
52	--	-.42	***-.75***
53	--	--	***-.59***
54	.42	-.50	***-.70***
55	.45	-.42	***-.46***
56	.41	***-.71***	-.58
57	--	***-.51***	--
58	--	***-.65***	-.59

Appendix C: Leadership Items Broken Down by Factor

Business Acumen

1. Quickly analyzes complex problems to find actionable, pragmatic solutions

2. Sees connections in data, events, trends, etc.

3. Consistently works against the right priorities

4. Creates effective and sustainable long-term strategies

5. Demonstrates a broad perspective when solving problems and making decisions

6. Integrates and simplifies complex or competing information for timely decision making

7. Looks ahead to reasonably anticipate business opportunities and obstacles

8. Takes the initiative to find ways to get better results

9. Actively seeks and takes advantage of ideas, best practices, and solutions developed elsewhere

10. Develops creative solutions and different ways of doing things that add value

11. Effectively filters through and identifies the best ideas to pursue

12. Fosters an environment of innovation and personally champions breakthrough ideas and initiatives

13. Supports reasonable risk taking and allows for some failures along the way

14. Demonstrates perseverance and resilience in the pursuit of goals

15. Confronts and works to resolve tough issues

16. Exhibits a “can do” attitude and willingness to take on significant challenges

17. Makes the right trade-offs and balances resources to deliver results

18. Plans ahead and articulates the organizational vision and resource requirements needed to deliver results for the future

19. Demonstrates the courage and conviction needed to drive large scale change initiatives

20. Demonstrates an awareness of his/her own strength and development areas

21. Seeks and acts on feedback from others regarding ways to improve his/her performance

22. Learns from experiences and can readily apply that knowledge to new situations

23. Seeks out learning opportunities and new experiences

24. Takes appropriate risks on people through the use of critical experiences and stretch assignments

25. Demonstrates a sense of passion, enjoyment, and pride about his/her work

26. Embraces and adapts well to change

27. Articulates a compelling vision, purpose, and direction that inspires others to follow

28. Demonstrates courage and optimism during times of crisis or change

29. Continually raises the bar for performance and helps others succeed

30. Keeps others informed so there are no unnecessary surprises

31. Anticipates controversial issues and questions and provides effective solutions

Inclusiveness

1. Treats all people with respect and fairness

2. Demonstrates sensitivity to differences when dealing with people from different cultural backgrounds and/or other differences

3. Demonstrates openness to and respect for others' opinions and points of view

4. Champions diversity of thought, style, and perspective

5. Demonstrates sensitivity and awareness of cross-cultural implications when conducting business or executing initiatives

6. Creates a work environment that helps people achieve a healthy balance between work and personal life

7. Fosters a positive and inclusive work environment where all people feel respected and valued for their contributions

8. Demonstrates a positive attitude in the workplace

9. Creates a work environment that makes work rewarding and enjoyable

10. Collaborates well with others to deliver results

11. Effectively listens to and understands what other people are saying

12. Builds relationships outside his/her sphere of influence to get things done

13. Operates effectively in matrix relationships across organizational boundaries

14. Creates a safe and trusting work environment that encourages openness and honest dialogue

15. Builds trusting relationships with multiple internal and external stakeholders for the broader good of the organization

16. Fosters an environment where people feel comfortable speaking with truth and candor

Integrity

1. Displays responsible and ethical behavior toward customers, suppliers, and others in the organization

2. Consistently follows the meaning and intent of company policies and standards (e.g., Code of Conduct)

3. Acts with integrity on a daily basis even when it is difficult to do so

4. Makes strategic decisions based on ethical considerations to ensure the long-term sustainability of the organization

5. Ensures systems, processes, and practices reinforce the PepsiCo values

6. Creates an ethical culture by encouraging and rewarding acting with integrity

7. Can be counted on to speak with truth and candor

8. Acts consistently with his/her words

9. Takes accountability for his/her own mistakes

Appendix D: Results of Supplemental Analysis Controlling for the Impact of Rater Gender

Table 1: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Adjustment on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Adjustment by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.730	.036	98.087	104.124	.000
Rater Gender	-.056	.086	119.354	-.654	.514
Adjustment	.000	.001	93.126	.229	.820
Adjustment x Purpose	-.002	.001	63.224	-2.175	.033

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 2: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Adjustment on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Adjustment by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.754	.035	95.741	107.489	.000
Rater Gender	.036	.086	113.966	.421	.675
Adjustment	-.001	.001	92.049	-.873	.385
Adjustment x Purpose	-.001	.001	70.074	-.586	.560

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 3: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Adjustment on aggregate of administrative ratings and developmental ratings of Integrity, and the Adjustment by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.977	.035	94.162	113.597	.000
Rater Gender	-.029	.088	109.578	-.332	.740
Adjustment	-.001	.001	92.786	-.433	.666
Adjustment x Purpose	-.002	.001	87.397	-1.917	.059

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 4: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Ambition on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Ambition by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.732	.036	97.438	104.134	.000
Rater Gender	-.029	.088	109.578	-.332	.740
Ambition	-.001	.001	94.658	-.691	.491
Ambition x Purpose	-.002	.001	69.629	-1.713	.091

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 5: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Ambition on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Ambition by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.755	.035	95.672	107.752	.000
Rater Gender	.035	.086	113.719	.414	.680
Ambition	-.001	.001	91.996	-.952	.344
Ambition x Purpose	-.001	.001	77.548	-.484	.630

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 6: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Ambition on aggregate of administrative ratings and developmental ratings of Integrity, and the Ambition by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.978	.035	93.726	114.163	.000
Rater Gender	-.028	.087	108.850	-.319	.750
Ambition	-.001	.001	94.629	-1.062	.291
Ambition x Purpose	-.003	.001	120.285	-2.148	.034

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 7: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Sociability on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Sociability by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.732	.036	97.535	103.973	.000
Rater Gender	-.061	.087	117.587	-.704	.483
Sociability	-.002	.001	90.552	-1.257	.212
Sociability x Purpose	-.003	.001	57.752	-2.532	.014

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 8: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Sociability on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Sociability by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.753	.035	95.832	107.160	.000
Rater Gender	.035	.086	112.522	.404	.687
Sociability	-.000	.001	89.227	-.265	.791
Sociability x Purpose	-.003	.001	59.303	-2.912	.005

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 9: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Sociability on aggregate of administrative ratings and developmental ratings of Integrity, and the Sociability by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.976	.035	93.630	113.402	.000
Rater Gender	-.036	.088	109.174	-.404	.687
Sociability	-.001	.001	89.817	-.773	.441
Sociability x Purpose	-.003	.001	83.006	-2.927	.004

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 10: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Interpersonal Sensitivity by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.737	.036	97.608	104.891	.000
Rater Gender	-.100	.089	123.270	-1.133	.259
Int. Sensitivity	.002	.001	104.626	1.382	.170
Int. Sensitivity x Purpose	-.000	.001	61.311	-.143	.887

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 11: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Interpersonal Sensitivity by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.754	.035	96.479	107.073	.000
Rater Gender	.028	.089	120.392	.319	.750
Int. Sensitivity	.000	.001	104.879	.352	.726
Int. Sensitivity x Purpose	.000	.001	70.039	.529	.598

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 12: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Interpersonal Sensitivity on aggregate of administrative ratings and developmental ratings of Integrity, and the Interpersonal Sensitivity by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.979	.035	94.697	113.626	.000
Rater Gender	-.048	.091	114.448	-.527	.599
Int. Sensitivity	.000	.001	102.485	.427	.670
Int. Sensitivity x Purpose	.001	.001	76.693	.483	.630

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 13: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Prudence on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Prudence by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.734	.036	99.608	103.591	.000
Rater Gender	-.069	.087	118.594	-.782	.436
Prudence	.001	.001	86.233	.654	.515
Prudence x Purpose	.001	.001	51.562	.750	.457

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 14: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Prudence on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Prudence by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.754	.035	97.378	106.999	.000
Rater Gender	.033	.086	114.843	.383	.702
Prudence	.000	.001	84.993	.130	.897
Prudence x Purpose	.001	.001	60.448	.495	.623

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 15: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Prudence on aggregate of administrative ratings and developmental ratings of Integrity, and the Prudence by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.980	.035	95.785	113.366	.000
Rater Gender	-.044	.088	110.251	-.496	.621
Prudence	.000	.001	85.823	.290	.772
Prudence x Purpose	.002	.001	70.424	1.222	.226

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 16: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Learning Approach (LA) on aggregate of administrative ratings and developmental ratings of Business Acumen, and the Learning Approach by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.732	.036	97.667	103.876	.000
Rater Gender	-.065	.087	117.366	-.753	.453
LA	-.001	.001	94.674	-.390	.698
LA x Purpose	-.002	.001	55.574	-1.668	-.101

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 17: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Learning Approach (LA) on aggregate of administrative ratings and developmental ratings of Inclusiveness, and the Learning Approach by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.753	.035	95.297	107.162	.000
Rater Gender	.031	.086	112.585	.361	.719
LA	.000	.001	93.826	.033	.974
LA x Purpose	-.002	.001	62.935	-2.005	.049

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.

Table 18: Regression coefficients, standard errors, degrees of freedom, critical *t*, and significance levels for the effect of rater gender and rater Learning Approach (LA) on aggregate of administrative ratings and developmental ratings of Integrity, and the Learning Approach by Purpose interaction.

Predictor	Estimate	s.e.	df	t	*p*
Intercept	3.978	.035	93.865	113.850	.000
Rater Gender	-.042	.088	109.549	-.473	.637
LA	.000	.001	92.636	.167	.868
LA x Purpose	-.002	.001	73.547	-1.724	.089

Note. Rater Gender was dummy coded such that 0 = Male and 1 = Female. Purpose was effect coded such that 1 = Administrative and -1 = Developmental.